

Online publizierte Arbeiten zur Linguistik

1/2016

Im Auftrag des Instituts für Deutsche Sprache
herausgegeben von Hardarik Blühdorn, Mechthild Elstermann und Annette Klosa

Vera Hildenbrandt, Annette Klosa (Hg.):

Lexikographische Prozesse bei Internetwörterbüchern

doi:10.14618/opal_01-2016



Institut für Deutsche Sprache
Postfach 10 16 21
68016 Mannheim
opal@ids-mannheim.de

Mitglied der Leibniz-Gemeinschaft



© 2016 IDS Mannheim – Alle Rechte vorbehalten

Das Werk einschließlich seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechts ist ohne Zustimmung der Copyright-Inhaber unzulässig und strafbar. Das zulässige Zitieren kleinerer Teile in einem eigenen selbstständigen Werk (§ 51 UrhG) erfordert stets die Angabe der Quelle (§ 63 UrhG) in einer geeigneten Form (§ 13 UrhG). Eine Verletzung des Urheberrechts kann Rechtsfolgen nach sich ziehen (§ 97 UrhG). Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen. Die zugänglichen Daten dürfen von den Nutzern also nur zu rein wissenschaftlichen Zwecken genutzt werden. Eine darüber hinausgehende Nutzung, gleich welcher Art, oder die Verarbeitung und Bearbeitung dieser Daten mit dem Zweck, sie anschließend selbst oder durch Dritte kommerziell zu nutzen, bedarf einer besonderen Genehmigung des IDS (Lizenz). Es ist nicht gestattet, Kopien der Textdateien auf externen Webservern zur Verfügung zu stellen oder Dritten auf sonstigem Wege zugänglich zu machen. Bei der Veröffentlichung von Forschungsergebnissen, in denen OPAL-Publikationen zitiert werden, bitten die Autoren und Herausgeber um eine entsprechende kollegiale Information an opal@ids-mannheim.de.

Inhalt

Vera Hildenbrandt, Annette Klosa

Einleitung 2

Melina Alexa, Anja Konopka, Lutz Wind

Gibt es einen medienspezifischen lexikografischen Prozess für das Online-Wörterbuch? –

Ein Werkstattbericht 8

Carole Tiberius, Tanneke Schoonheim

The *Algemeen Nederlands Woordenboek* (ANW) and its Lexicographical Process 20

Annette Klosa

Der lexikographische Prozess im Projekt *elexiko* 29

Sandra Denzer, Franziska Horn

Die ersten Schritte eines Langzeitprojekts.

Das Digitale Familiennamenwörterbuch Deutschlands 39

Christian M. Meyer, Iryna Gurevych

Der lexikographische Prozess im deutschen Wiktionary 56

Rufus H. Gouws

Aspekte des lexikographischen Prozesses in Print- und Onlinewörterbüchern 70

Vera Hildenbrandt, Annette Klosa

Lexikographische Prozesse bei Internetwörterbüchern

Einleitung

Während lexikographische Prozesse,¹ die zur Publikation gedruckter Wörterbücher führen, bereits seit einigen Jahrzehnten im Fokus der Wörterbuchforschung stehen² und die von Wiegand dafür unterschiedenen Phasen der Vorbereitung, der Datenbeschaffung, der Datenaufbereitung, der Datenauswertung und der Satz- und Druckvorbereitung (Wiegand et al. 2010, S. 8ff.) mittlerweile als etabliert betrachtet werden dürfen,³ steht die Diskussion und Beschreibung lexikographischer Prozesse von Internetwörterbüchern noch in den Anfängen. Zwar besteht kein Zweifel daran, dass sich lexikographische Prozesse bei der Publikation von Internetwörterbüchern anders gestalten als bei Printwörterbüchern, doch die Fragen, inwiefern sie dies tun, welchen Einfluss die neuen Möglichkeiten der Datengewinnung aus elektronischen Textkorpora auf die Prozesse haben, wie Bearbeitungsteilwortschätze auszuwählen sind, wie verschiedene Fassungen zu versionieren und zu archivieren sind und wie sich schließlich die Änderungen der lexikographischen Prozesse auf die Nutzer auswirken, ob und wie die Nutzer in diese Prozesse einbezogen werden können, sind noch nicht ausführlich beantwortet. Diese und andere Fragen waren daher Gegenstand des vierten Arbeitstreffens des wissenschaftlichen Netzwerks „Internetlexikografie“, das am 22. und 23. November 2012 an der Universität Trier stattfand und vom Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften/Trier Center for Digital Humanities organisiert wurde.

Im Rahmen des Arbeitstreffens wurden zunächst kurze und vertiefte Einblicke in die lexikographischen Prozesse verschiedener Wörterbücher gegeben. Mit dem Trierer Wörterbuchnetz (www.woerterbuchnetz.de) wurde eine lexikographische Online-Publikation präsentiert, die ausschließlich aus retrodigitalisierten, überwiegend historisch ausgerichteten Wörterbüchern besteht. Auch zu den lexikalischen Ressourcen des Digitalen Wörterbuchs der deutschen Sprache (www.dwds.de) gehören retrodigitalisierte Wörterbücher. Darüber hinaus aber entwickelt das DWDS ein korpusbasiertes Digitales Lexikalisches System, das die Wörter des deutschen Wortschatzes in ihren Eigenschaften beschreibt.⁴ Das Mittelhochdeutsche Wörterbuch (www.mhdwb-online.de) steht zwischen gedruckt und online erscheinendem Nachschlagewerk; alle Phasen seiner Erstellung werden computerunterstützt durchgeführt, konzeptionell führend ist allerdings das gedruckte Wörterbuch, die nahezu zeitgleich erscheinende Internetpublikation wird davon abgeleitet. Auch Duden online (www.duden.de, vgl. den Beitrag von Melina Alexa, Anja Konopka und Lutz Wind) entsteht aus derselben Datenbasis wie die gedruckten Wörterbücher des Dudenverlags, allerdings gibt es medienspezifische Unterschiede im Her-

¹ Bewusst sei hier der Plural verwendet, da es – dies hat sich nicht zuletzt in den Diskussionen des Arbeitstreffens gezeigt – den lexikographischen Prozess schlechthin nicht gibt, da dessen Ausgestaltung von verschiedenen Faktoren – u.a. etwa von Typ und Entstehungsweise des Wörterbuchs – abhängt.

² Einen knappen Überblick über den Stand der Forschung gibt Klosa (2013, S. 519).

³ Vgl. etwa die Beiträge in diesem Arbeitsbericht, die sich nahezu ausnahmslos auf Wiegand beziehen.

⁴ Vgl. dazu Geyken (2014).

stellungsprozess der online und der im Print erscheinenden Ressourcen. Mit dem Algemeen Nederlands Woordenboek (<http://anw.inl.nl/search>, vgl. den Beitrag von Carole Tiberius und Tanneke Schoonheim) und *elexiko* (www.owid.de/wb/elexiko/start.html, vgl. den Beitrag von Annette Klosa) entstehen von vornherein digital konzipierte Online-Wörterbücher der niederländischen respektive deutschen Gegenwartssprache, mit dem Digitalen Familiennamenwörterbuch Deutschlands (vgl. den Beitrag von Sandra Denzer und Franziska Horn) ein „digital born“-Spezialwörterbuch deutscher Familiennamen. Ebenfalls ausschließlich für die Online-Publikation entwickelt ist das allgemeine, mehrsprachige Wörterbuch Wiktionary (<http://de.wiktionary.org/wiki/Wiktionary:Hauptseite>, vgl. den Beitrag von Christian M. Meyer und Iryna Gurevych), das im Gegensatz zu den bereits angeführten Wörterbüchern allerdings nicht redaktionell, sondern gemeinschaftlich erarbeitet wird.

Ob nun retrodigitalisiert, für den Druck konzipiert und dann online publiziert oder ausschließlich digital entworfen und im Internet veröffentlicht, ob abgeschlossenes oder Ausbauwörterbuch, ob redaktionell als Verlagsprodukt, von einer wissenschaftlichen Einrichtung erarbeitetes oder kollaborativ erstelltes Werk – die Einblicke verdeutlichten, dass sich die lexikographischen Prozesse dieser Internetwörterbücher oder -informationssysteme von denen ausschließlich gedruckt erscheinender Wörterbücher unterscheiden. Zugleich aber demonstrierten sie, dass die lexikographischen Prozesse abhängig sind von der Art des Internetwörterbuchs, dass also etwa der eines gemeinschaftlich erstellten Wörterbuchs anders ist als der eines redaktionell betreuten. Und schließlich zeigten sie die Wörterbuchredakteure auf der Suche nach einer adäquaten Beschreibung der lexikographischen Prozesse. Den praktischen Einblicken folgte ein theoretischer Beitrag, der auf methodologischer und metalexikographischer Ebene die Notwendigkeit einer allgemeinen Lexikographietheorie diskutierte, die als Basis dienen kann für die Definition verschiedener Arten von lexikographischen Prozessen von Print- und Internetwörterbüchern (vgl. den Aufsatz von Rufus Gouws).

Die Auseinandersetzung mit dem lexikographischen Prozess wurde fortgesetzt in drei Arbeitsgruppen, die sich mit Auswahlkriterien, Umsetzung und Problemen von Bearbeitungsteilwortschätzen, mit Archivierung und Versionierung und mit dem korpusbasierten Vorgehen bei der Erweiterung bestehender lexikographischer Ressourcen beschäftigten. Auf die eingangs formulierten Fragen wurden in den Diskussionsrunden und Arbeitsgruppen die folgenden Ergebnisse gefunden bzw. weiterführende Fragen identifiziert.⁵

1. Inwiefern gestalten sich lexikographische Prozesse bei der Publikation von Internetwörterbüchern anders als bei Printwörterbüchern?

Dass beim Versuch der Beschreibung lexikographischer Prozesse von Internetwörterbüchern immer wieder Bezug genommen wird auf das für die Entstehung gedruckter Wörterbücher entwickelte 5-Phasen-Modell Wiegands,⁶ verdeutlicht, wie sehr die Internetlexikographie sich hier noch in einer Zeit des Umbruchs befindet. Besonders redaktionell erstellte Internetwörterbücher berufen sich zunächst auf diese Phasierung, erweitern sie allerdings um eine Phase

⁵ Die im Folgenden gegebene Zusammenfassung der Diskussion basiert auf den Protokollen der Hilfskräfte des Trierer Kompetenzzentrums: Sandra Behrend, Semjon Borchert, Katharina Dietz, Matthias Schneider, Aline Stang und René Tobner sei an dieser Stelle herzlich gedankt.

⁶ Vgl. die Beiträge von Annette Klosa, Christian Meyer/Iryna Gurevych, Carole Tiberius/Tanneke Schoonheim und Rufus Gouws.

der Digitalisierung⁷ oder Computerisierung,⁸ in der die nötige Software bereitgestellt wird und Analysewerkzeuge zum Einsatz kommen. Die konkrete Umsetzung der einzelnen Phasen allerdings unterscheidet sich in Abhängigkeit vom entstehenden Wörterbuchtyp zum Teil fundamental von denen gedruckter Wörterbücher. Dies gilt in besonderer Weise für die letzte Phase in diesem Prozess, die Publikationsphase, die medienspezifisch anders abläuft, weniger auf ein abgeschlossenes als vielmehr auf ein offenes und dynamisches Produkt abzielt, das in regelmäßigen Abständen aktualisiert werden kann oder sich – wie Wiktionary – in einem beständigen Aktualisierungsvorgang befindet. Auf kollaborativ erarbeitete Wörterbücher scheinen denn auch die aus Printwörterbüchern bekannten Phasen nur in sehr begrenztem Maße anwendbar zu sein. Für Wiktionary wird daher ein neues, auf eine Vorbereitungs- und eine Bearbeitungsphase reduziertes Modell des lexikographischen Prozesses vorgeschlagen.⁹ Dabei werden diese Phasen – das wiederum ist allen im Rahmen des Arbeitstreffens vorgestellten Wörterbüchern gemeinsam – iterativ durchlaufen. Charakteristisch für lexikographische Prozesse bei Internetwörterbüchern ist also ein zirkulärer, nicht-linearer Verlauf, an dem im Vergleich zu Printwörterbüchern nicht überwiegend Lexikographen, sondern auch Computerlinguisten und Informatiker beteiligt sind und in den auch die potenziellen Nutzer verstärkt einbezogen werden können und müssen, sollten sie denn nicht ohnehin wie im Falle von Wiktionary gleichzeitig auch Wörterbuchautoren sein. Voraussetzungen für die produktive Gestaltung des lexikographischen Prozesses und das effiziente Einbeziehen des Nutzers sind ein vollständiges Durch- und immer wieder Überdenken dieses Prozesses und seiner Transparenz.

2. Wie sind Bearbeitungsteilwortschätze auszuwählen?

Die Nutzung von Internetwörterbüchern kann bereits gewährleistet sein, bevor sie vollständig bearbeitet sind. Dies wird u.a. durch die Arbeit in und die Publikation von Bearbeitungsteilwortschätzen ermöglicht, deren Auswahl in der Vorbereitungsphase des lexikographischen Prozesses mitgeplant werden sollte. Für die Arbeit in Bearbeitungsteilwortschätzen sprechen nicht nur mediale, sondern auch redaktionelle Gründe. Sie dient der Rationalisierung der lexikographischen Arbeit, erlaubt die systematische und umfassende Beschreibung vergleichender Wortschatzeinheiten und die Nutzung spezieller Expertise für bestimmte Wortschatzbereiche wie zum Beispiel etymologische Angaben oder Fachwortschätze. Allerdings bringt die Arbeit in Bearbeitungsteilwortschätzen auch Probleme mit sich. So entstehen etwa Reibungsverluste, wenn die auf Alphabetstrecken oder nach Wortarten verteilte Bearbeitung die konsistente Vernetzung der Stichwörter erschwert oder die Arbeit auf verschiedene Autorengruppen verteilt wird. Darüber hinaus ist es nicht ausgeschlossen, dass die Auswahl und Bearbeitung von Bearbeitungsteilwortschätzen an den Interessen der Nutzer vorbeigeht, da sie von projektinternen Gegebenheiten gesteuert wird. Soll die Bearbeitung in Bearbeitungsteilwortschätzen gelingen, müssen Lexikographen regelmäßig ihre Methoden überdenken und gemeinsam mit den Benutzern immer wieder eine reflexive Phase einschieben. Darüber hinaus sollte ein verständliches und erweiterbares Instruktionsbuch für eine gute Planung, Disziplin, Flexibilität, Kontrolle, Motivation und Transparenz sorgen.

⁷ Vgl. den Beitrag von Carole Tiberius und Tanneke Schoonheim.

⁸ Vgl. den Beitrag von Annette Klosa.

⁹ Vgl. den Beitrag von Christian Meyer und Iryna Gurevych.

3. Welchen Einfluss haben die neuen Möglichkeiten der Datengewinnung aus elektronischen Textkorpora auf die Prozesse?

Transparenz ist auch gefordert beim korpusbasierten Vorgehen in der lexikographischen Arbeit. Zur Beantwortung der Frage, welchen Einfluss die neuen Möglichkeiten der Datengewinnung aus elektronischen Korpora auf die lexikographischen Prozesse haben, wurden verschiedene Thesen aufgestellt:

These 1: Der Nutzen des Korpus und der Korpusrecherche hängt vom konkreten lexikographischen Ziel ab.

These 2: Ein Korpus ist nur so gut wie: a) die Abfragesprache, b) die Metadaten, c) die Größe eines sichtbaren Textausschnitts.

These 3: Gute Werkzeuge filtern und sortieren die Daten und machen sie handlich für die lexikographische Arbeit. Sie „übersehen“ aber alles, was selten ist.

These 4: „There is no data like more data.“

These 5: Das, was ins Wörterbuch aufgenommen wird, entspricht nicht dem, was im Korpus steht; die sprachliche Wirklichkeit steht also der lexikographischen Bearbeitung entgegen.

Aus diesen Thesen wurden für die Arbeit mit Korpusdaten im Wörterbuchprozess einige Forderungen abgeleitet: Korpustechnologen sollten nicht nur Statistiken und vorgefilterte Daten anbieten, sondern auch Data Mining-Funktionen wie z.B. Verteilungsdiagramme bereitstellen. Auch sollten sie automatisierte Mechanismen zur Kontrolle der sich verändernden Korpuszusammensetzung zur Verfügung stellen. Metadaten der Korpora sollten im Dialog mit Lexikographen erstellt und abfragbar gemacht werden. Lexikographen müssen im Umgang mit den Korpusdaten die Möglichkeit erhalten, die Daten zu annotieren. Dazu ist es jedoch zum einen erforderlich, dass sie in der Lage sind, über das Korpus hinauszublicken und zum anderen die statistisch vorgefilterten Daten kritisch analysieren zu können. Nur auf diese Weise können sie die Nutzer für die unterschiedliche Qualität der Datenarten sensibilisieren und geeignete Recherchetechniken in Wörterbuch- und Korpusdaten vermitteln. Den Nutzern sollte es in jedem Fall ermöglicht werden, die Sprachwirklichkeit weitgehend ungefiltert rezipieren zu können, d.h. neben dem Wörterbuch sollte das Korpus selbst in Gänze als eigenständige Publikation für verschiedene Anwendungen zur Verfügung stehen.

4. Wie sind verschiedene Fassungen eines Internetwörterbuchs zu versionieren und zu archivieren?

Auch Versionierung und Archivierung sind Teil der lexikographischen Prozesse und müssen bereits bei der Projektplanung bedacht werden. In Bezug auf die Versionierung besteht grundsätzlich kein Zweifel daran, dass die Daten permanent im jeweils aktuellen Zustand präsentiert werden und dass transparente Versionsmetadaten zur Verfügung stehen sollten, sodass für den Nutzer jederzeit ersichtlich ist, in welcher Wörterbuchversion er nachschlägt. Eine Timeline könnte hier gegebenenfalls für einen schnellen Überblick sorgen. Auch in Bezug auf die Archivierung gibt es eindeutige Forderungen. Die Wörterbuchdaten und -metadaten soll-

ten möglichst unabhängig von den technischen Bedingungen und Gegebenheiten in gut dokumentierten, offenen, standardisierten Formaten vorgehalten werden. Dabei sollten nicht nur Text-, sondern auch andere Medienformate wie beispielsweise Bilder und Karten bedacht werden. Die Daten sollten persistent zitier- und über Persistent Identifier adressierbar sein, wobei frühzeitig geklärt werden sollte, wie feingranular die Persistent Identifier sein können und sollen.

Aber auch offene Fragen und Probleme wurden in der Diskussion identifiziert. Ungeklärt ist etwa, ob Nutzer ein Interesse an früheren Versionen haben. Und auch die langfristige Vorhaltung von Infrastrukturen ist noch immer eine Herausforderung. Zwar ist die Notwendigkeit von Ressourcenkuration und -pflege, von Standardisierung und Archivierung zwischenzeitlich im Denken digital arbeitender Wissenschaftler verankert und wird auch von den Forschungsförderern gefordert,¹⁰ doch stehen für diesen wichtigen Teil digitaler wissenschaftlicher Tätigkeiten nach wie vor keine ausreichenden Förderformate zur Verfügung. Als andauernde Aufgabe ist in diesem Kontext die Ausbildung eines öffentlichen Bewusstseins für die neue digitale Kultur zu sehen.

5. Wie wirken sich Änderungen lexikographischer Prozesse auf Nutzer aus? Wie können Nutzer in diese Prozesse einbezogen werden?

Die potenziellen Nutzer von Internetwörterbüchern gehören zu dieser Öffentlichkeit und sollten verstärkt und von Beginn an in die lexikographischen Prozesse einbezogen werden. Der Umfang des Einbezogenwerdens und Einbezogenwerdenkönnens sowie des Nutzereinflusses richtet sich dabei natürlich nach dem Typ des Wörterbuchs und wird bei einem kollaborativ entstehenden Crowd-Sourcing-Wörterbuch wie Wiktionary größer sein als bei einem redaktionell erarbeiteten Wörterbuch wie *lexiko*. Einigkeit besteht jedoch darin, dass Nutzer bereits in der Planungs- oder Vorbereitungsphase des lexikographischen Prozesses eine wesentliche Rolle spielen sollten. Nach Möglichkeit sollten die verschiedenen Benutzergruppen identifiziert und die Funktionen des Wörterbuchs an ihren Bedürfnissen ausgerichtet, also unterschiedliche, unter Umständen auch interaktiv einzurichtende Nutzerprofile vorgesehen werden. Auch sollten die Nutzer jederzeit Rückmeldung geben können, was letztlich auch Auswirkungen auf die weitere lexikographische Arbeit haben kann.¹¹ Wichtig sind also zu jedem Zeitpunkt des lexikographischen Prozesses die Kommunikation mit den Nutzern, Offenheit und Transparenz.

Wie allerdings diese Transparenz geschaffen, für die Nutzer dargestellt werden kann, ist noch eine ebenso offene Frage wie die zukünftigen Nutzerbedürfnisse ungeklärt sind. Die Notwendigkeit empirischer, auch auf eine mittel- bis langfristige Perspektive gerichteter Nutzerstudien ist also offenkundig. Zugleich aber sollten Lexikographen, Wissenschaftler und versierte Wörterbuchnutzer angehalten werden, Nutzungsmöglichkeiten zu offerieren und zu erklären, um so Potenziale aufzuzeigen und Interesse und Bedarf zu wecken.

Insgesamt führte das Arbeitstreffen in aller Deutlichkeit vor Augen, wie sehr lexikographische Prozesse (in Abhängigkeit vom Wörterbuchtyp) im Wandel begriffen sind, wie sehr sich die zur Bewältigung der Prozesse notwendigen Kompetenzen und damit Berufsgruppen ausweiten, wie sehr die Nutzer von Beginn an einbezogen werden sollten und wie wichtig Transpa-

¹⁰ Vgl. etwa die DFG-Praxisregeln „Digitalisierung“ (Deutsche Forschungsgemeinschaft (Hg.) 2013).

¹¹ Zur Rolle des *simultaneous feedback* im lexikographischen Prozess vgl. den Beitrag von Rufus Gouws.

renz und Kommunikation in allen Prozessphasen sind. Es zeigte aber auch, wie viele offene Fragen es noch gibt und wie wichtig eine theoretische Reflexion der lexikographischen Prozesse bei Internetwörterbüchern ist.

Literatur

- Deutsche Forschungsgemeinschaft (Hg.) (2013): DFG-Praxisregeln „Digitalisierung“. DFG Vordruck 12.151-02/13. Bonn: Deutsche Forschungsgemeinschaft. www.dfg.de/formulare/12_151/12_151_de.pdf (Stand: 30.7.2015).
- Geyken, Alexander (2014): Methoden bei der Wörterbuchplanung in Zeiten der Internetlexikographie. In: *Lexicographica* 30, S. 77-112.
- Klosa, Annette (2013): The lexicographical process (with special focus on online dictionaries). In: Gouws, Rufus Hjalmar/Heid, Ulrich/Schweickard, Wolfgang/Wiegand, Herbert Ernst (Hg.): *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography.* (= *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK)* 5.4). Berlin/Boston: de Gruyter, S. 517-524.
- Wiegand, Herbert Ernst/Beißwenger, Michael/Gouws, Rufus H./Kammerer, Matthias/Storrer, Angelika/Wolski, Werner (Hg.) (2010): *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research.* Bd. 1: A-C. Berlin/New York: de Gruyter.

Melina Alexa, Anja Konopka, Lutz Wind

Gibt es einen medienspezifischen lexikografischen Prozess für das Online-Wörterbuch? – Ein Werkstattbericht

1. Einleitung

Im Jahr 2010 hatte sich der Verlag Bibliographisches Institut GmbH dazu entschieden, ein Internetwörterbuch zu erarbeiten und es unter dem Name *Duden online* auf der Verlagswebseite www.duden.de zu veröffentlichen. *Duden online* wurde als ein eigenständiges Wörterbuch geplant, das sich aus Wörterbuchdaten verschiedener Provenienz speist. Diese Daten wurden – und werden – in einer verlagseigenen Datenbank, dem Duden-Wissensnetz,¹ abgelegt. Am 11. Mai 2011 ging das Internetwörterbuch online. Schon aus dieser kurzen Skizzierung der Entstehungsgeschichte von *Duden online* lässt sich erkennen, dass der Herstellung von *Duden online* und den gedruckten Wörterbüchern eine gemeinsame Basis zugrunde liegt: Die Wörterbuchdaten für die Printwörterbücher und das Online-Wörterbuch werden in derselben Datenbank geführt – eine wichtige Gemeinsamkeit, doch natürlich gibt es im weiteren Verlauf des Herstellungsprozesses auch medienspezifische Unterschiede. Sind sie groß genug, um von einem „medienspezifischen Prozess“ für *Duden online* zu sprechen? Um eine Antwort auf diese Frage zu finden, sollen im Folgenden die lexikografischen Prozesse von Print- und Onlinewörterbuch einander gegenübergestellt werden.² Dabei ist zunächst zu überlegen, unter welchen Voraussetzungen ein gedrucktes Wörterbuch einerseits und ein Onlinewörterbuch andererseits entsteht, welche allgemeinen Erwartungen an die beiden Medienformen gestellt werden und welche medienspezifischen Besonderheiten es gibt in Bezug auf die Länge des Produktzyklus, das Platzangebot und die Erwartungen der Leser/Benutzer.³

2. Medienunabhängige Content-Bearbeitung im „Duden – Wissensnetz deutsche Sprache“

Bevor wir die beiden Prozesse der Print- und Onlineproduktion von Wörterbüchern bzw. der redaktionellen Verarbeitung von Wörterbuchinhalten ausführlich vorstellen, sei zunächst die gemeinsame Grundlage der beiden Prozesse kurz vorgestellt: das sogenannte *Duden – Wissensnetz deutsche Sprache*, kurz *Wissensnetz* genannt. Das Wissensnetz ist die seit dem Jahr 2000 im Verlag BI GmbH aufgebaute elektronische Ressource, in der die wichtigsten Wörterbuchsubstanzen des Dudenverlags bearbeitet und verwaltet werden. Basierend auf dem Prin-

¹ Alexa (2011, S. 61-70).

² Neben gedruckten Wörterbüchern und dem Internetwörterbuch *Duden online* publiziert Duden auch Wörterbücher in anderen Medienarten, z.B. als App, als Download oder als CD-ROM. Da deren lexikografische Datengrundlage, Publikationsrhythmus und Herstellungsprozess in weiten Teilen dem des zugrunde liegenden gedruckten Wörterbuchs folgen, wird auf diese nicht näher eingegangen.

³ Aus Gründen der Lesbarkeit wird das generische Maskulinum verwendet. Hiermit sind ausdrücklich Frauen und Männer gemeint.

zip einer objektorientierten Datenbank wurden die bereits vorhandenen lexikografischen Duden-Substanzen im Wissensnetz zusammengeführt und kontinuierlich bearbeitet und erweitert.⁴

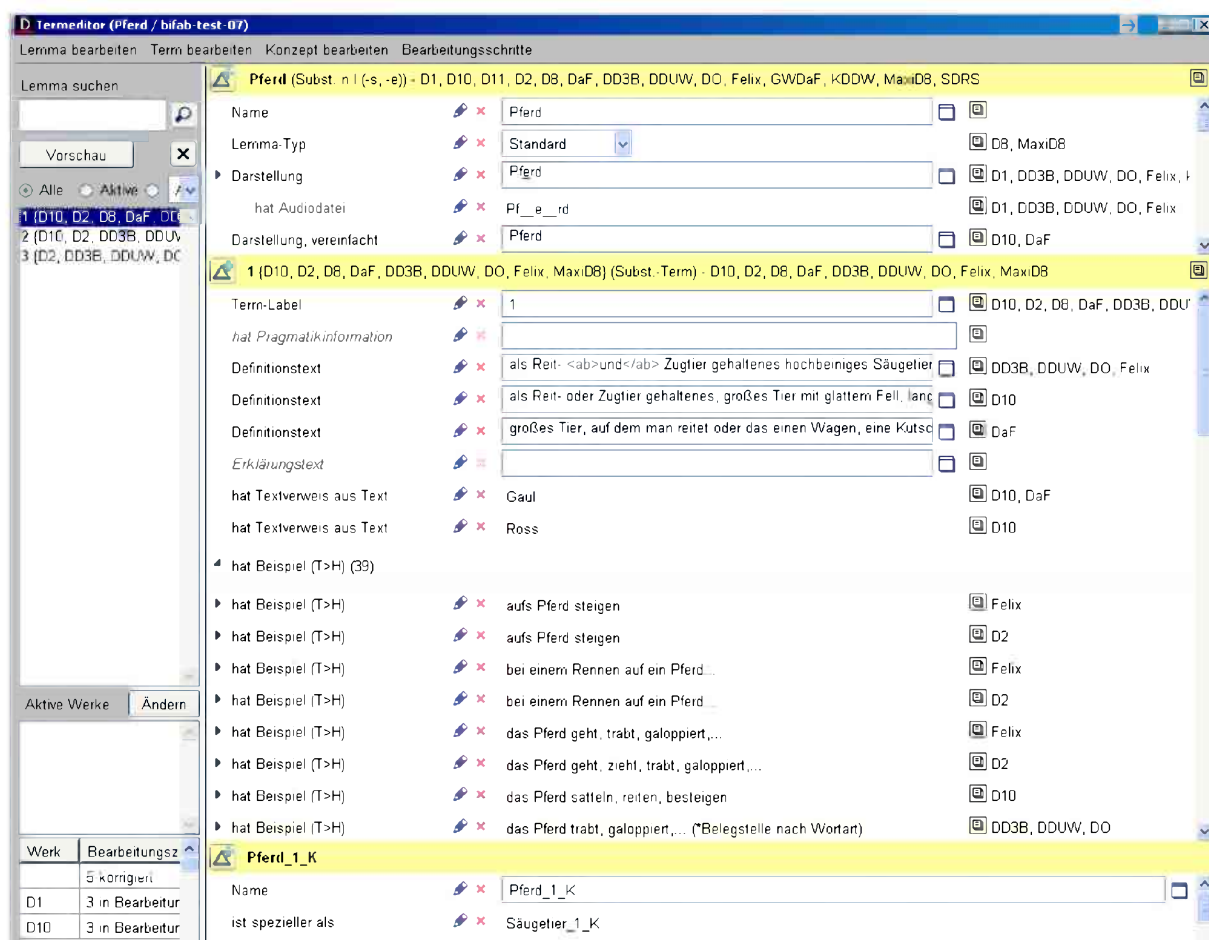


Abb. 1: Auszug der Eigenschaften des Lemmas *Pferd* im Wissensnetz

Das Ziel zu Projektbeginn war es, auf den sich abzeichnenden Rückgang des Printmarktes zu reagieren und den Duden-Content für die Zukunft möglichst medienneutral und bearbeitungs- wie kosteneffizient vorzuhalten. Im Laufe des Projekts wurden die Einzelsubstanzen der wichtigsten Duden-Wörterbücher in einem objektorientierten Wissensrepräsentationssystem – dem Wissensnetz – zusammengeführt und dort weitergepflegt und ausgebaut. Aus der so entstandenen Datenbank werden heute, wie ursprünglich geplant, nicht nur Printwörterbücher produziert, sondern auch die Daten für elektronische Offline- und Online-Produkte zum Nachschlagen oder Korrigieren und für Contentlizenzgeschäfte im Geschäftskundenbereich exportiert.

Die aus der zentralen Datenablage und Datenbearbeitung resultierende Redundanzminimierung war ein wichtiger Aspekt des Wissensnetzprojekts, ein weiterer Vorteil ergab sich aus der besonderen Art der Datenhaltung: eine semantisch-ontologische Vernetzung linguistischer Daten, basierend auf einem Begriff-Instanzen-Modell.

⁴ Alexa et al. (2002, S. 1-8).

In diesem System ist ein Lemma die konkrete Instanz eines allgemeineren Begriffs (z.B. das Lemma *Pferd* ist eine Instanz der als Begriff modellierten grammatischen Wortklasse „Substantiv Neutrum“).

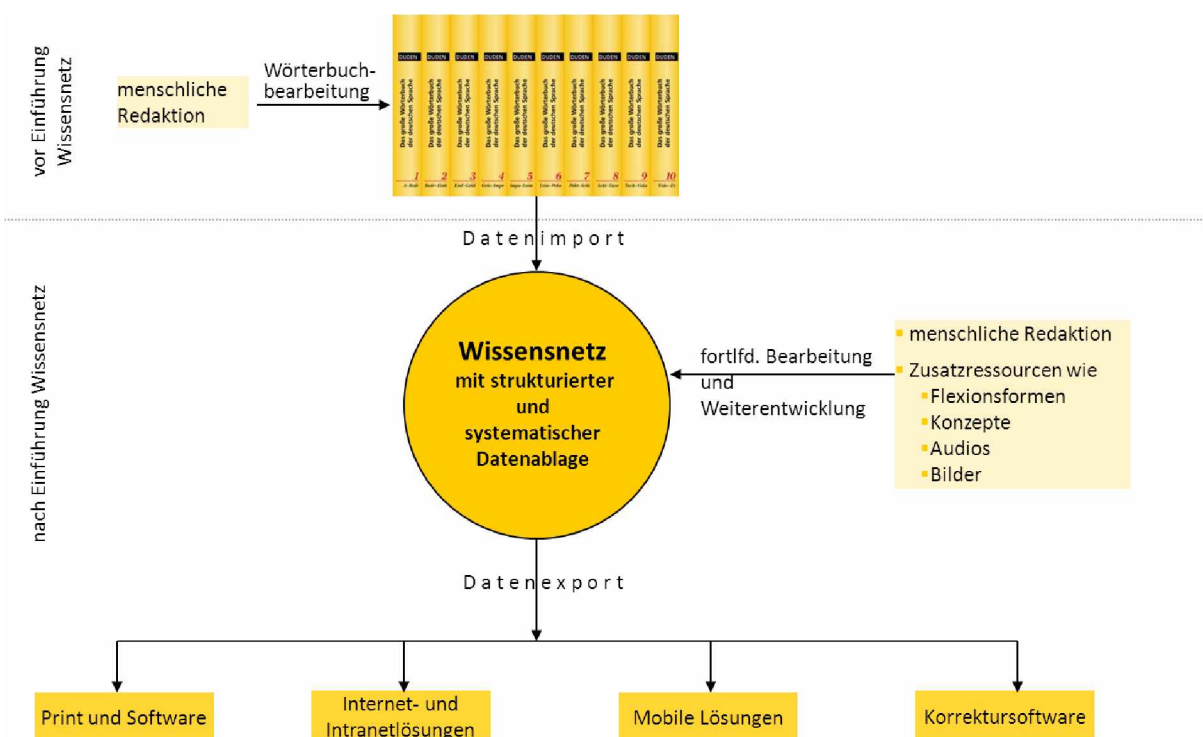


Abb. 2: Das Wissensnetz als zentrales, medienneutrales Redaktionswerkzeug

Die Instanzen tragen in diesem Repräsentationssystem bestimmte Eigenschaften, mit denen die lexikografischen und linguistischen Informationen dem Lemma zugewiesen werden, nämlich Attribute und Relationen. Mithilfe von zweistelligen Relationen lassen sich Lemmata oder weitere Instanzen im Netz, wie etwa Verwendungsbeispiele oder idiomatische Wendungen, untereinander verknüpfen. Man drückt auf diese Weise der Wissensrepräsentation z.B. aus, dass ein Lemma die orthografische Variante eines anderen oder sinnverwandt/synonym zu einem anderen Lemma ist, bestimmte Verwendungsbeispiele hat oder in einem oder mehreren bestimmten Duden-Werken vorkommt. Mittels der Attribute (auch als einstellige Relationen aufzufassen) werden einem Lemma individuelle Eigenschaften zugeordnet, wie beispielsweise die, dass das Wort *Pferd* nicht trennbar ist und sich etymologisch aus mittelhochdeutsch *pfert* ableitet. Diese Strukturierung der lexikografischen wie linguistischen Informationen erlaubt den einfachen systematischen Zugriff auf diese Informationen, sei es zur Suche in den Daten oder zur algorithmischen Bearbeitung oder Zusammenstellung großer Datenmengen nach bestimmten Kriterien. Daneben steht die effiziente lexikografische Artikelbearbeitung: Ein Lemma wird zentral für alle Duden-Werke bearbeitet, in denen es verwendet wird. Neben der bereits angesprochenen Redundanzminimierung ist dadurch eine werkübergreifende Konsistenz einfacher sicherzustellen und es können bereits vorhandene Informationen aus anderen Werken übernommen werden – ein Aspekt, der gerade bei der Zusammenstellung der Artikel für *Duden online* eine wesentliche Rolle gespielt hat. In Abb. 1 ist ein Ausschnitt des Artikels *Pferd* im Wissensnetz dargestellt, der an einer Stelle für insgesamt 14 Duden-Wörterbücher die zentrale Informationsdichte verdeutlicht.

Neben diesen klassisch lexikografischen Daten wurden kontinuierlich auch solche in das Wissensnetz aufgenommen, die in der Regel keine Verwendung in Print-Nachschlagewerken finden. Dazu gehören Bilder, Aussprachedateien, die vollständigen flektierten Formen der drei Hauptwortarten Substantiv, Verb und Adjektiv sowie Frequenzangaben zur Vorkommenshäufigkeit im Dudenkorpus. Durch diese Daten ist man flexibler in der Lage, neue sprachtechnologische Produkte zu konzipieren oder Dienstleistungen anzubieten.

In Abb. 2 sind schematisch der initiale Aufbau, der kontinuierliche Ausbau und die Verwendung des Wissensnetzes als Produktquelle dargestellt.

Ein wesentlicher Schritt in den Prozessketten um das Wissensnetz ist die Exportprogrammierung. Der Export im Wissensnetz sorgt dafür, dass die wie oben beschriebenen vernetzten Daten werk- bzw. verwendungsbezogen aus der Datenbank herausgeschrieben werden. Dazu läuft der Export durch das Wissensnetz und sammelt die produktspezifisch verknüpften Informationseinheiten zusammen. Das Resultat sind Daten in XML, die je nach Produkt weiterverarbeitet werden. Bei einer Printproduktion erfolgt die Übergabe einer XML-Strecke zur Verarbeitung in weiteren Ressourcen, bis sie letztlich im Satzsystem angekommen sind (siehe die Skizzierung der Workflows im folgenden Kapitel). Für *Duden online* erfolgt die Bearbeitung direkt von den Wissensnetz- bzw. für den Content Verantwortlichen der Duden-Sprachtechnologie. Über die XML-Daten laufen danach verschiedene Skripte, die z.B. die korrekte Verweisverlinkung zwischen den Artikeln sicherstellen, und über XSL-Transformationen wird schließlich die XML-Strecke in einzelne HTML-Dateien aufgesplittet und dem Online-Content-Management-System zur Aktualisierung der Daten auf duden.de übergeben.

3. Medienspezifische Merkmale und ihre Auswirkungen auf den lexikografischen Prozess

3.1 Der lexikografische Prozess bei Print-Wörterbüchern

3.1.1 Produktzyklus/Aktualisierung

Nach dem Blick auf das Verbindende der Duden-Wörterbücher, die gemeinsame Vorhaltung der Wörterbuchdaten im Wissensnetz, soll nun auf die medienspezifischen Unterschiede eingegangen werden. Sie resultieren vor allem aus abweichenden Produktzyklen beim Print- bzw. Onlinewörterbuch, einem unterschiedlichen Platzangebot und einer jeweils anders gearteten Rolle des Benutzers. Beginnen wir mit einer genaueren Betrachtung des lexikografischen Prozesses in Bezug auf die Printwörterbücher und wenden uns dort zunächst der Frage zu, wie sich der printtypische Produktzyklus auf den lexikografischen Prozess auswirkt. Gewöhnlich findet eine grundlegende Überarbeitung und Aktualisierung des Wörterbuchinhalts im Rahmen einer Neuauflage statt. Dies ist beim Duden ca. alle vier bis fünf Jahre der Fall. Ein gedrucktes Wörterbuch kann also nur mit relativ langer Verzögerung auf aktuelle Wortschatz- und Sprachentwicklungen reagieren. Für eine Neubearbeitung wird dann in der Regel das ganze Wörterbuch von A bis Z gelesen und überarbeitet; alle Anfragen, Anmerkungen, Verbesserungsvorschläge und Korrekturhinweise, die zu diesem Werk in der Zwischenzeit eingegangen sind, werden geprüft und – falls sie sich als begründet erweisen – eingearbeitet. Da die Redaktion dies bereits im Voraus sorgfältig plant, indem sie Umfang und Zeitbedarf der jeweiligen Arbeitspakete (Neuaufnahmen, Überarbeitung grammatischer, etymologischer, phonologischer und stilistischer Informationen, Ergänzung von Bedeutungen) berechnet, ist eine konsistente Bearbeitung des Wörterbuchs über die gesamte A-Z-Strecke hinweg gewährleistet.

3.2.1 Das Platzangebot

Lexikografen stehen bei der Bearbeitung eines im Druck erscheinenden Wörterbuchs immer vor dem gleichen Problem: Platzknappheit. Die Einhaltung des geplanten Umfangs ist ein zentrales Kriterium sowohl bei der Auswahl der aufzunehmenden Lemmata als auch bei der Ausgestaltung der Wörterbuchartikel. Der Wunsch, möglichst viele Informationen auf knappem Raum zu zeigen, lässt sich nur realisieren, indem die Informationen hoch verdichtet werden, z.B. durch Verwendung von Abkürzungen, Ersetzung des Stichworts durch einen Ersetzungsstrich („-“), Andeutung des Flexionsparadigmas in der Grammatikangabe durch die Beschränkung auf Nominativ Singular, Genitiv Singular und Nominativ Plural, restriktive Auswahl der Beispiele, Verzeichnung idiomatischer Wendungen an einer einzigen Stelle (d.h. beim ersten Substantiv dieser Wendung). Die starke Verdichtung der lexikografischen Informationen macht das Lesen gedruckter Wörterbücher oft mühsam; ungeschulten und unerfahrenen Benutzern gelingt es häufig nicht, die gesuchte Information zu finden, obwohl sie im Wörterbuch steht, und den Inhalt eines Wörterbuchartikels vollständig zu verstehen. Die beiden folgenden Beispiele zu *Krempe* und *Eichhörnchen* sind dem Duden-Universalwörterbuch (Dudenredaktion (Hg.) 2011) entnommen:

Krem|pe, die; -, -n [aus dem Niederd., eigtl. = die Gekrümmte, verw. mit †Kranpf]: *Hutkrempe*: eine schmale, breite K.; die K. nach oben biegen, herunterdrücken.

Abb. 3: Artikel *Krempe* aus dem Universalwörterbuch

Eich|hörn|chen, das: *kletterndes, rotbraunes bis schwarzbraunes Nagetier mit langem, buschigem Schwanz*: R mühsam [er]nährt sich das E. (ugs. scherzh.; *die Ausführung [dieses Vorhabens] ist langwierig, schwierig u. nur in kleinen Schritten möglich*).

Abb. 4: Artikel *Eichhörnchen* aus dem Universalwörterbuch

3.1.3 Die Rolle der Benutzer

Da das Medium Buch kein tages-, wochen- oder monatsaktuelles Medium ist, kann der Leser zwar Verbesserungsvorschläge machen und Änderungswünsche mitteilen, muss sich aber relativ lange gedulden, bis er überprüfen kann, ob sein Wortvorschlag tatsächlich berücksichtigt oder seine kritische Anmerkung umgesetzt wurde. Lob und Kritik zu den gedruckten Werken erreichen die Redaktion vorwiegend auf zwei Wegen: Über die telefonische Sprachberatung⁵ und über Leserzuschriften, die per Post oder per E-Mail an den Verlag gesendet werden. Eine endgültige Antwort, ob z.B. das vorgeschlagene Wort in eines oder sogar mehrere Wörterbücher aufgenommen wird, darf der Benutzer aber nicht sofort erwarten: Über die Aufnahme von Wörtern wird eben erst im Rahmen einer Neubearbeitung der Werke entschieden, wenn abzuschätzen ist, wie viel Platz für Neuaufnahmen vorhanden ist.

⁵ Vgl. dazu Scholze-Stubenrecht (2002, S. 225-248) und Pellengahr, Carsten (2008, S. 24-25).

3.1.4 Prozessstationen (Workflow)

Zuletzt sei noch ein Blick darauf geworfen, wie der Weg von der Projektidee „Wir machen ein (gedrucktes) Wörterbuch!“ bzw. „Wir machen eine Neuauflage des Wörterbuchs ...!“ bis zum gedruckten Buch verläuft. Genannt seien hier nur die wichtigsten Stationen: Nach der Konzepterstellung und anschließender Produktentscheidung beginnen die inhaltlichen Arbeiten, die Erstellung des Contents. Es folgen die technische Bearbeitung der Daten (Konvertierung, Erstellen einer Satzvorlage, Datenexport in das Satzssystem etc.), die Gestaltung des Layouts, bis schließlich am Ende das Imprimatur steht.

3.2 Der lexikografische Prozess bei Online-Wörterbüchern

3.2.1 Produktzyklus/Aktualisierung

Wie schon in der Einleitung erwähnt wurde, ist *Duden online* ein eigenes Werk und nicht die digitale Version eines bereits in gedruckter Form vorliegenden Wörterbuchs. Somit besitzt es auch einen eigenen Aktualisierungsrythmus. Um den Erwartungen der Benutzer an ein Internetwörterbuch gerecht zu werden, aber auch aus Gründen der Suchmaschinenoptimierung – Websites, die für Nutzer wertvolle Inhalte anbieten und die regelmäßig überarbeitet werden, werden von Suchmaschinen in der Ergebnisliste höher gerankt – wird der Inhalt von *Duden online* in relativ kurzen Abständen aktualisiert, d.h. es werden neue Wörter aufgenommen, gegebenenfalls Fehler korrigiert, Bedeutungen ergänzt und vieles mehr (vgl. dazu auch die Kap. 3.2.3 und 3.2.4). Das heißt nicht, dass damit alte lexikografische Prinzipien zur Aufnahme neuer Wörter über Bord geworfen werden: Nicht jede Gelegenheitsbildung wird aufgenommen und ein Wort, das gewissermaßen „über Nacht“ in aller Munde ist, morgen vielleicht aber schon in Vergessenheit geraten sein könnte, wird erst einmal auf einen „Warteplatz“ verwiesen. Aber gute Wortvorschläge, neue Bedeutungen und Ähnliches finden auf diese Weise viel schneller den Weg ins Wörterbuch, als es sonst der Fall ist.

Gleichzeitig bedeutet dies natürlich auch, dass bei einer Aktualisierung keine A-Z-Bearbeitung durchgeführt wird, sondern punktuelle Veränderungen vorgenommen werden.

3.2.2 Platzangebot

Während der Lexikograf bei der Erarbeitung/Bearbeitung eines gedruckten Wörterbuchs immer den Umfang im Auge behalten muss, befindet er sich bei einem Onlinewörterbuch in der komfortablen Lage, dass er im Prinzip unbegrenzt Platz zur Verfügung hat. Dieser Umstand wird in mehrfacher Hinsicht genutzt:

Zum einen werden Informationen, die auch in einem gedruckten Wörterbuch vorhanden sind, dort aber in sehr komprimierter Form gezeigt werden, in *Duden online* expliziter, ausführlicher und damit verständlicher präsentiert: So werden z.B. die einzelnen Informationseinheiten eines Wörterbuchs explizit benannt und in bestimmte Rubriken (Rechtschreibung, Grammatik, Aussprache, Herkunft, Bedeutung etc.) eingeordnet. Außerdem wird nahezu gänzlich auf die Verwendung von Abkürzungen verzichtet und auch das Lemma selbst wird innerhalb des Artikels stets ausgeschrieben:

Krempe, die

Wortart: Substantiv, feminin

Häufigkeit: ■■■■

Rechtschreibung ⓘ

[↑ Nach oben](#)

Worttrennung:

Krem|pe

Bedeutung ⓘ

[↑ Nach oben](#)

Hutkrempe

Beispiele

- eine schmale, breite Krempe
- die Krempe nach oben biegen, herunterdrücken

Synonyme ⓘ zu Krempe

[↑ Nach oben](#)

Hutrand

Aussprache ⓘ

[↑ Nach oben](#)

Betonung:

Krempe

Lautschrift:

[*ˈkʁɛmpə*]

Herkunft ⓘ

[↑ Nach oben](#)

aus dem Niederdeutschen, eigentlich = die Gekrümmte, verwandt mit ↑*Krampf*

Abb. 5: Ausschnitt aus dem Artikel *Krempe* in *Duden online*

Zum anderen werden auch Grammatikkommentare ausformulierter präsentiert:

Bei Substantiven tritt an die Stelle der aus Artikel, Genitiv- und Pluralendung bestehenden Grammatikangabe eine explizit formulierte grammatische Information; man vergleiche einen für gedruckte Wörterbücher typischen Eintrag wie z.B. bei *Espresso*

Espresso; der u. die, -, -[s] od. ...ssi <aber: 3 Espresso> mit der entsprechenden Angabe in *Duden online*:



der Espresso; Genitiv: des Espresso[s], Plural: die espressos oder espressi <aber: drei Espresso>

Abb. 6: Grammatikangabe bei einem Substantiv in *Duden online*

Auch grammatische Angaben z.B. bei Verben sind im Onlinewörterbuch sehr viel weniger verdichtet als im gedruckten wie ein Vergleich zwischen Duden-Universalwörterbuch

<st. V.; hat> :
und *Duden online* zeigt:



starkes Verb; Perfektbildung mit »hat«

Abb. 7: Grammatikangabe bei einem Verb in *Duden online*

Schließlich ermöglicht das große Raumangebot auch die Einbindung von Inhalten, die in gedruckten Bedeutungs- und Rechtschreibwörterbüchern in der Regel nicht enthalten sind, wie etwa Flexionsparadigmen für Substantive, Verben und Adjektive, Häufigkeitsangaben, Kollokationen, Bilder und Sounddateien zur Aussprache (vgl. Abb. 8):

Eichhörnchen, das

Wortart: Substantiv, Neutrum

Häufigkeit: ■■■■

Rechtschreibung ⓘ

⬆ Nach oben

Worttrennung:

Eich|hörn|chen

Bedeutung ⓘ

⬆ Nach oben

kletterndes, rotbraunes bis schwarzbraunes Nagetier mit langem, buschigem Schwanz

Wendungen, Redensarten, Sprichwörter

mühsam [er]nährt sich das Eichhörnchen (*umgangssprachlich scherzhaft*; die Ausführung [dieses Vorhabens] ist langwierig, schwierig und nur in kleinen Schritten möglich)



© Joanna Redesiuk - Fotolia.com

Synonyme ⓘ zu Eichhörnchen

⬆ Nach oben

(landschaftlich) **Eichhase**, **Eichkater**, **Eichkätzchen**, **Eichkatze**; (besonders Zoologie) **Eichhorn**

Aussprache ⓘ

⬆ Nach oben

Betonung:

Eichhörnchen 🔊

Grammatik ⓘ

⬆ Nach oben

	Singular	Plural
Nominativ	das Eichhörnchen	die Eichhörnchen
Genitiv	des Eichhörnchens	der Eichhörnchen
Dativ	dem Eichhörnchen	den Eichhörnchen
Akkusativ	das Eichhörnchen	die Eichhörnchen

Abb. 8: Artikel *Eichhörnchen* in *Duden online*: Mehr Platz als in Print-Wörterbüchern ermöglicht die Einbindung von Bildern, Aussprachdateien und Flexionsparadigmen

3.2.3 Die Rolle des Benutzers

Wie bereits in Kapitel 3.2.1 angedeutet wurde, hat der Benutzer einen viel direkteren Einfluss auf den Inhalt des Onlinewörterbuchs als bei gedruckten Wörterbüchern. Die Anregungen der Benutzer haben in mehrfacher Hinsicht Auswirkungen auf *Duden online*: Konkrete Ergänzungs- oder Korrekturhinweise führen kurzfristig zu punktuellen Ergänzungen wie beispielsweise Neuaufnahmen oder Korrekturen, die dann beim nächsten Datenexport berücksichtigt werden. Es gibt jedoch auch große Bereiche des Wörterbuchs betreffende systematische Änderungen, die – teils auf Anregung der Benutzer hin, teils durch aktive Benutzerbefragung – seitens des Redaktionsteams vorgenommen werden. Ein Beispiel dafür ist die Onlinebefragung, die ein Jahr nach dem Go-Live von *Duden online* stattfand. Ein wichtiges Ergebnis dieser Umfrage war, dass sich viele Benutzer einen Ausbau der Synonyme und der Flexionsformenangaben wünschten; Bildern hingegen wurde kein besonders großer Wert beigemessen. Dieses Ergebnis wurde bei der Priorisierung des geplanten Ausbaus von *Duden online* berücksichtigt und führte zu einem systematischen Ausbau sowohl des Synonymen- als auch der Flexionsformenbestandes.⁶

3.2.4 Suchmaschinenoptimierung

Wenn auch die Wörterbuchinhalte die Website von *Duden online* dominieren, so dient die Seite dennoch nicht allein Informationszwecken: Auf www.duden.de wird auch für Produkte des Verlages geworben.⁷ Dies hängt mit der hinter *Duden online* stehenden Geschäftsidee zusammen, dass eine starke Frequentierung der *Duden-online*-Sites zu hohen Verkaufszahlen der dort beworbenen Produkte führt. Insofern liegt es im Interesse des Verlages, dass die Website in Suchergebnislisten hoch gerankt wird und möglichst viele Zugriffe hat. Wie lässt sich das erreichen? Ein Faktor wurde bereits angesprochen: Die Betreiber von Suchmaschinen „belohnen“ Websites, die gut gepflegt sind und wertvolle Informationen anbieten, mit einem besseren Ranking. Auch werden starke Marken begünstigt. Ein weiterer wichtiger Faktor ist, dass der im Internet eingegebene Suchbegriff genau so auf der Website steht, und zwar am besten nicht nur ein Mal, sondern mehrere Male. Steht das gesuchte Wort mehrmals auf einer Seite, ist dies für die Suchmaschinen ein Indikator dafür, dass die Seite eine hohe Relevanz besitzt. Außerdem wird eine hohe Anzahl von Links „honoriert“. Was bedeutet das konkret für *Duden online*?

Zum einen folgt daraus, dass die Inhalte des Wörterbuchs so aufbereitet sein müssen, dass der Benutzer das gesuchte Wort oder die gesuchte Wendung möglichst auf Anhieb finden kann. Dies ist nicht selbstverständlich, denn viele Laienbenutzer geben Wörter in flektierter Form ein und wundern sich dann, dass sie keinen Treffer erhalten. Auch Mehrwortausdrücke wie *zum einen ... zum anderen* lassen sich unter Umständen nicht unmittelbar finden, da Wörterbücher in der Regel keine Mehrwortlemmata ansetzen. Für die lexikografische Bearbeitung folgt daraus, dass möglichst viele Flexionsparadigmen generiert und in *Duden online* integriert werden müssen, damit der Benutzer auch flektierte Formen finden kann (und natürlich sind sie auch eine wertvolle Information für jeden, der sich über die richtige Flexion eines Wortes im Unklaren ist und nicht eine Grammatik zur Hand nehmen möchte). Zudem wurden viele Wendungen, besonders wenn sie orthografisch schwierig waren, in *Duden online* als

⁶ Rautmann (2013, S. 49-62).

⁷ Diese Angaben beziehen sich auf die Websitegestaltung zum Zeitpunkt des Vortrags im November 2012. Mittlerweile wird auf duden.de auch Werbung für andere Unternehmen geschaltet.

Stichwörter aufgenommen, z.B. das oben bereits erwähnte *zum einen ... zum anderen* oder *bis auf Weiteres*. Zum anderen wurden die Inhalte stark miteinander verlinkt, was natürlich wiederum nicht nur im Interesse der Suchoptimierung, sondern auch der Benutzer ist. So sind etwa fast alle Synonyme als Links realisiert, Verweise natürlich auch.

3.3 Prozessesstationen (Workflow)

Der Workflow des Onlinewörterbuchs unterscheidet sich auf den ersten Blick von dem eines Printwörterbuchs in mehrfacher Hinsicht: Der Aktualisierungsrhythmus ist wesentlich kürzer, anstatt von A-Z-Bearbeitungen finden in der Regel thematische Bearbeitungen statt und es differieren durch das Entfallen des Mediumträgers Papier, an dessen Stelle das Internet tritt, die Prozessbeteiligten. Dennoch gibt es auch viele Gemeinsamkeiten, vor allem, wenn man die verschiedenen Prozessesstationen betrachtet: Um eine Aktualisierung anzustoßen, bedarf es keiner Produktentscheidung, sondern regelmäßige Aktualisierungen sind Bestandteil des Konzepts und wurden von Beginn an geplant. Die Überarbeitung des Inhalts erfolgt – wie beim Printprozess – im Wissensnetz und wird vom *Duden-online*-Team sowie von Lexikografen der Dudenredaktion vorgenommen. Es schließt sich auch hier die technische Bearbeitung der Daten an, diesmal nicht für den Export in ein Satzprogramm, sondern in ein Content-Management-System, es folgen der Proof des Probeexports und der abschließende Export ins Livesystem. Geändert hat sich im Ablauf vor allem der Zeitpunkt des Korrekturlesens: Erfolgt das Korrekturlesen bei Printwörterbüchern erst, nachdem die erste Fahne erzeugt wurde, also de facto nach Abschluss der Manuskriptarbeiten, geschieht dies im Online-Workflow meist direkt im Anschluss an das Verfassen oder Bearbeiten eines Artikels.

4. Fazit

Die diesem Artikel überschriebene Frage, ob es einen medienspezifischen lexikografischen Prozess für das Online-Wörterbuch gibt, lässt sich aus unserer Sicht und somit aus der Perspektive des Dudenverlags hinsichtlich des Wörterbuch-Contents unter duden.de insgesamt mit einem eingeschränkten „Ja“ beantworten: Die Besonderheit des Mediums und die Anforderungen der Nutzer haben deutliche Auswirkung auf die Konzepte, die erstellt werden, und diese beeinflussen ihrerseits natürlich erheblich den Prozess und die lexikografische Arbeit. Der Umstand, dass beim Online-Medium keine Platzbeschränkung herrscht und immer wieder azyklisch zu Printprojekten und punktuell Änderungen oder Ergänzungen vorgenommen werden können, ist ein Aspekt. Ein weiterer ergibt sich aufgrund der besonderen Spielregeln im Online-Geschäft, die teilweise zur besonderen Aufarbeitung der Inhalte führen, nämlich derart, dass Suchmaschinen die Webseiten möglichst prominent in ihren Suchergebnislisten anzeigen. Auch kann der Kunde bzw. Benutzer des Online-Wörterbuchs mehr oder minder direkt und kurzfristig Einfluss auf dessen Gestaltung und Inhalte nehmen, sei es über die Teilnahme an Umfragen oder durch ein individuelles Feedback. Last but not least unterscheidet sich der Prozessablauf auch in Bezug auf die Qualitätssicherung, die für die Online-Inhalte ein anderes Vorgehen erfordert und während regelmäßiger und frequenterer Content-Aktualisierungen nicht sinnvollerweise über Komplettlesungen wie im Print-Prozess gelöst werden kann.

Allerdings gibt es auch viele Gemeinsamkeiten: Die Kernbeteiligten bei der Erstellung von Print- und Onlineprodukten sowie ihre Funktionen in diesen Prozessen sind im Wesentlichen identisch, beginnend bei der zentralen Datenhaltung und redaktionellen Bearbeitung bis hin zum Export des Contents und dessen Weitergabe zur Fertigstellung/Onlineschaltung des Produkts.

Allerdings sind die Kompetenzen der Einzelnen erweitert worden und die Affinität der Beteiligten zur Online-Publikation und Online-Nutzung ist eine Voraussetzung für den Erfolg eines Onlinewörterbuchs.

5. Literatur

- Alexa, Melina (2011): Modellierung eines semantischen Wissensnetzes für lexikographische Anwendungen am Beispiel der Duden-Ontologie. In: Klosa, Annette/Müller-Spitzer, Carolin (Hg.): Datenmodellierung für Internetwörterbücher. 1. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“. (= OPAL – Online publizierte Arbeiten zur Linguistik 2/2011). Mannheim, S. 61-70. <http://pub.ids-mannheim.de/laufend/opal/pdf/opal2011-2.pdf>.
- Alexa, Melina/Kreissig, Bernd/Liepert, Martina/Reichenberger, Klaus/Rostek, Lothar/Rautmann, Karin/Scholze-Stubenrecht, Werner/Stoye, Sabine (2002): The Duden ontology: an integrated representation of lexical and ontological information. In: Proceedings of LREC-2002, OntoLex-Workshop 2002: Ontologies and Lexical Knowledge Bases, 27th May 2002, Las Palmas, Canary Islands, Spain. Las Palmas, S. 1-8.
- Dudenredaktion (Hg.) (2011): Duden – Deutsches Universalwörterbuch. 7., überarb. u. erw. Aufl. Red. Bearb. Werner Stolze-Stubenrecht. Mannheim.
- Pellengahr, Carsten (2008): Der heiße Draht nach Mannheim. In: Tribüne 1, S. 24-25.
- Rautmann, Karin (2013): Duden online und seine Nutzer. In: Abel, Andrea/Klosa, Annette (Hg.): Der Nutzerbeitrag im Wörterbuchprozess. 3. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“. (= OPAL – Online publizierte Arbeiten zur Linguistik 4/2014). Mannheim, S. 49-62. <http://pub.ids-mannheim.de/laufend/opal/pdf/opal2014-4.pdf>.
- Scholze-Stubenrecht, Werner (2002): „Die Auswahl der Einträge ist äußerst beliebig.“ Warum Jagdherr und Pokémon nicht im Duden stehen. In: Sprachwissenschaft 27, 2, S. 225-248.

Carole Tiberius, Tanneke Schoonheim

The *Algemeen Nederlands Woordenboek* (ANW) and its Lexicographical Process

1. The *Algemeen Nederlands Woordenboek* (ANW)

The *Algemeen Nederlands Woordenboek* (ANW, Dictionary of Contemporary Dutch) is an online, corpus-based, scholarly dictionary of contemporary standard Dutch in the Netherlands and in Flanders, describing the Dutch vocabulary from 1970 onwards. It is one of the main projects of the Leiden Instituut voor Nederlandse Lexicologie (INL, Institute of Dutch Lexicology). As well as being an ‘online dictionary’ through which a range of users can explore the Dutch vocabulary, the ANW is also a linguistic data resource from which especially language professionals can extract data necessary for their research. The project focuses on the general vocabulary of written Dutch and it provides semasiological and onomasiological access to the dictionary.

The ANW can be characterised as an online dictionary under construction (called an *Ausbauwörterbuch* by Schröder 1997). The project started in 2001 and a first version of the full dictionary should be completed at the end of 2017. In December 2009 a demo version was launched and new dictionary articles are being added on a regular basis (with an average of four updates per year).¹ In this paper, we describe the lexicographical process of the ANW from its conception through to its online publication.

2. The lexicographical process of the ANW

As Klosa (2013, p. 519) notes, publishing an online dictionary under construction affects the lexicographical process. The three phases that can theoretically be distinguished in any lexicographical process – planning, writing, and producing (Landau 1984, p. 227) – merge in an online dictionary under construction. This also applies to the ANW where we see an overlap in the planning and writing phase and in the writing and production phase.

The result is a more complex lexicographical process, which Klosa (2013) calls the ‘computer-lexicographical process’. Adapting work by Wiegand (1999, p. 233), she distinguishes six phases in the computer-lexicographical process, i.e.,

- Preparation
- Data acquisition
- Computerisation
- Data processing
- Data analysis
- Preparation for online release

Our description of the lexicographical process of the ANW will follow these six phases.

¹ For more information on the ANW see Schoonheim/Tempelaars (2010) and references on the ANW website: http://anw.inl.nl/show?page=help_publicaties.

2.1 The phase of preparation

As for any other dictionary, the computer-lexicographical process for an online dictionary starts with a phase of preparation and planning. This involves an organisational plan that contains details on finance, work flow, schedule, and personnel. For a digitally processed and online displayed dictionary the plan needs to take into account the deployment of computational linguists, corpus linguists, and software engineers, whom are very important to guide the old-fashioned lexicographical process into the digital age. It is not enough to try to convert manual procedures to automatic procedures, but it is necessary to rethink the whole process in order to see how the computer can be used to make it more effective and efficient.

The organisational plan of the ANW included funding for the project until 2018. At the beginning of the project, it seemed reasonable to employ only one full time software engineer / computational linguist to the project and to use the rest of the money for lexicographical skilled employees. If asked to make this choice again with present day knowledge, we would opt for less lexicographical staff and more computational staff – at least two, but may be even three –, to be able to reduce the manual part of the lexicographical process as much as possible by using smart tools and new lexicographical techniques.

Besides the organisational plan, a conceptual design plan needs to be written during this first phase, with information on the content of the dictionary. The data structure needs to be defined, sample entries must be written, the lexicographer's manual needs to be compiled (although this will be revised several times during the writing phase), and the corpus has to be designed. The conceptual plan should also contain information on the design of the online application as well as the required functionalities (e.g. supported search strategies). Finally, a plan should be written on how to monitor the users of the dictionary.

The conceptual plan of the ANW was written in the early years of this millennium. The data structure was defined as a collaborative effort between lexicographers and the software engineer, sample entries were written and the initial type templates for the semagrams² (Moerdijk 2008) were developed. A first version of the lexicographer's manual was compiled and internal reports were written on topics such as terminology, proper names, abbreviations, and the treatment of collocations.

As the ANW is a corpus-based dictionary, plans were made for compiling a balanced corpus of contemporary standard Dutch, including both material from the Netherlands as well as from Flanders.³ During this period, an in-house dictionary writing system was also developed. Furthermore, two documents were written in the period 2007-2009, describing the functional and technical requirements of the web application of the ANW dictionary.

2.2 The phase of data acquisition

The basis for an online dictionary (as for any other dictionary) includes primary sources, i.e., an electronic text corpus, and secondary sources, such as other paper or electronic dictionaries, as well as grammar books, lexical databases, and so on.

² Semagrams are an innovative aspect of the ANW dictionary. They are systematic representations of the knowledge associated with a word in a frame of slots and fillers.

³ http://anw.inl.nl/show?page=help_anwcorpus

Being a corpus-based dictionary, the primary source of the ANW is a corpus. In the case of the ANW, this corpus was compiled specifically for the project and consists of several subcorpora: literary texts (20%), newspaper material (40%), domain-specific texts (35%), and a subcorpus with texts containing neologisms (5%). Originally, corpus compilation was completed in 2004, when a corpus size of just over 100 million tokens⁴ was reached. However, we are now in the process of moving from a static corpus to a dynamic corpus by adding new material to the various subcorpora starting with newspaper material.

There are also a number of secondary sources which the ANW lexicographers have at their disposal when editing dictionary entries. These are lexicographical sources, e.g. the *Groot Woordenboek van de Nederlandse Taal* by Van Dale, the *Woordenboek der Nederlandsche Taal* (WNT),⁵ the *Oxford English Dictionary* (OED),⁶ and *elexiko*,⁷ linguistic sources as the *Morfologisch Handboek* and the *Algemene Nederlandse Spraakkunst* (ANS), and finally documentary sources such as Wikipedia and Google. The lexicographers can access both primary and secondary sources from within the dictionary writing system.

2.3 The phase of computerisation

The main phase of computerisation of the ANW took place roughly between 2004 and 2009. During this period, the technical equipment has been set up. The selected corpus material has been annotated with tagging and lemmatisation software (cf. De Does/Van der Voort van der Kleij 2002) developed at INL and has been loaded into a corpus query system (Tiberius/Kilgariff 2009). Originally an in-house corpus system was used, but as of 2007, the ANW project uses the Sketch Engine (Kilgariff et al. 2004). In the same period, the Dictionary Writing System was developed and the first draft of the database structure for presenting the dictionary articles was defined. The database structure (2.3.1), the dictionary writing system (2.3.2), and the corpus query system (2.3.3) are continuously being improved based on new insights both from the lexicographers as well as from the computational linguists and software engineers.

2.3.1 Specifying the database structure

The database structure of the ANW dictionary has been specified by the software engineer, based on the wishes and needs of the lexicographers. The ANW uses XML. An XML schema was defined in early 2006 and has gradually been fine-tuned over the years. On the basis of this XML schema, the user interface of the ANW editor is generated. The ANW data is stored in a MySQL database together with metadata, such as the author of the article, an overview of the elements which have been completed and the status of the article (e.g. being edited, ready to go to editor in chief, online, etc.). For the online application, the data is exported from the MySQL database and converted to a structure which makes it possible to search more efficiently through the data.

⁴ A corpus of one hundred million tokens is considered to be large enough for describing the normal use of a language (cf. Hanks 2002, p. 157).

⁵ <http://gtb.inl.nl>

⁶ www.oed.com

⁷ www.owid.de/wb/elexiko/start.html

2.3.2 The lexicographic workstation and the ANW-editor

The ANW uses a Dictionary Writing System which was designed specifically for the project. It consists of two parts, the lexicographic workstation and the ANW-editor. The workstation is basically a menu bar which appears at the top of the screen and allows the lexicographers to invoke various tools and resources facilitating the editing process from raw material to the neat dictionary article. Important elements are of course a lemma list from which the lexicographer chooses a lemma and opens it for editing. In addition, the lexicographic workstation contains links to the secondary sources mentioned above including links to electronically available specialist literature, other dictionaries, internal documents, as well as a list of editorial guidelines.

When the lexicographer chooses a lemma from the lemma list and opens it for editing, the ANW-editor is started up. Similar to the lexicographic workstation, this editor has been designed specifically for the project (Niestadt 2009). The editor adopts an explorer approach, meaning that elements from the article structure can be opened and closed at will, which is beneficial to the general overview during the editing process, especially considering the rich information structure of the ANW.

The basic information structure of the ANW contains ten main categories, which in turn are subdivided into one or more subcategories, depending on the complexity of the subject. For instance, the main category 'Lemma' contains the subcategories 'Lemma form', 'Variants', and 'Lemma type'. In a number of cases the choice of a specific element in the main category determines the subcategories to be shown. If a lexicographer chooses the option 'substantief' ('noun') as the value for 'syntactic category type', he/she is shown the data sheet for nouns to complete, whereas if he/she would have chosen 'werkwoord' ('verb'), the data sheet for verbs would have opened up. In the editorial process, cross-references, and other internal links in the dictionary are automatically checked for consistency.

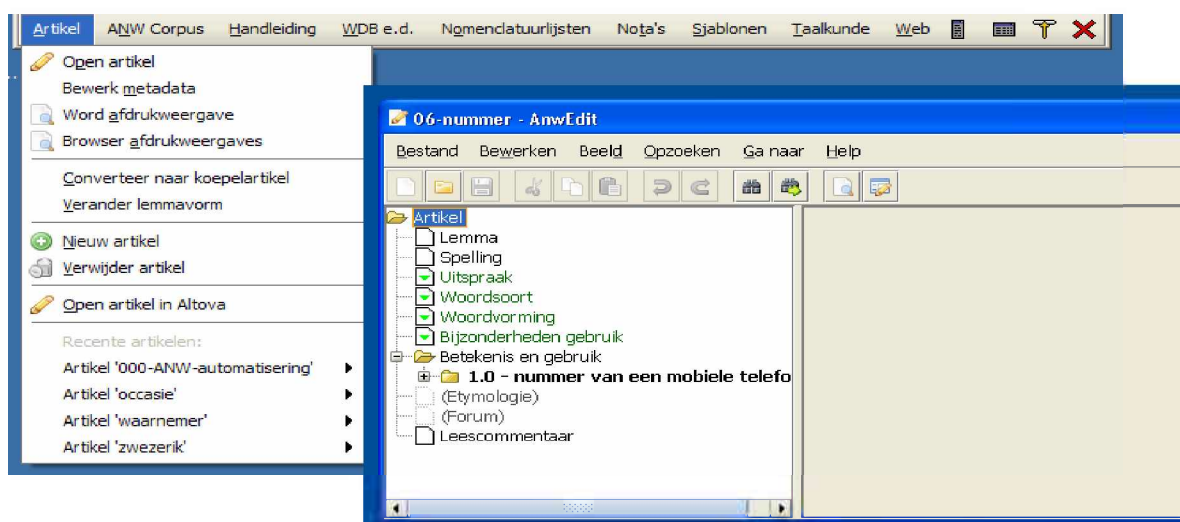


Figure 1: View of the ANW Dictionary Writing System

2.3.3 The Sketch Engine

During the editing process the lexicographer has access to the ANW corpus in the Sketch Engine (Kilgarriff et al. 2004). The Sketch Engine allows the lexicographer to sort the material in various ways and to deduce information on the usage of the lemmas. Combinations and collocations can also easily be found, using options such as sort by context. Another useful feature is TickBox Lexicography (Kilgarriff et al. 2010), which has been set up for the ANW in such a way that it makes it possible to import not only the relevant collocations directly into the editor, but also copies the corresponding examples and source information immediately in the right place in the article structure. An equally important asset of the Sketch Engine is GDEX (Good Dictionary EXamples), which assists the lexicographer in selecting examples for the dictionary that best illustrate the different meanings, collocations, and possibilities of use of the word (Kilgarriff et al. 2008). We are currently (2015) experimenting with including automatically selected example sentences in the dictionary.

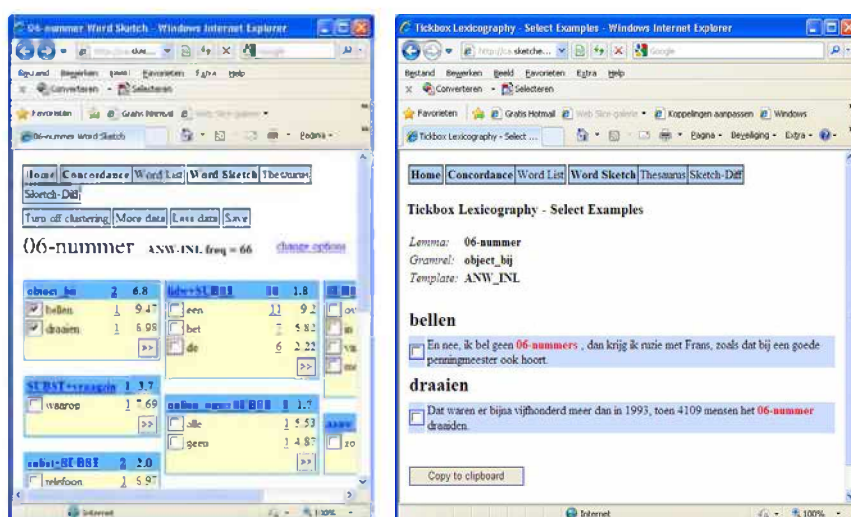


Figure 2: The Sketch Engine talks to the ANW editor

2.4 The phase of data processing

After the data has been acquired and after it has been automatically prepared for use, it can be processed for further analysis. In the initial phase of data processing the preliminary lemma list has been compiled and candidate entries have been selected (2.4.1). A Sketch Grammar was written (2.4.2) and an inventory of lexicographic data that can be obtained by automatic data acquisition was compiled (2.4.3). This inventory is continuously being updated. Klosa (2013, p. 520) also includes the collection, recording, and tagging for online use of external sources such as audio files or videos in the phase of data processing. In the ANW the collection of multimedia is a continuous task, which is performed by the lexicographic assistants for the words that go online in the next update.

2.4.1 Compilation of the lemma list, choice of entry candidates

On the basis of the ANW corpus a lemma list was derived. This corpus-based lemma list was compared to authorised word files such as the *Woordenlijst der Nederlandse Taal*, the *Referentiebestand Nederlands*, and the *Referentiebestand Belgisch-Nederlands*. This resulted in a balanced selection of the entry candidates for the ANW.

Not all words in the dictionary will necessarily get a complete lexical-semantic description, because the meaning of many derived words is clear from the word elements involved. It is, however, the intention of the ANW to provide all selected words at least with information on spelling, pronunciation, abbreviation, flexion, and morphology. Compounds and derivations without full lexical-semantic treatment are mentioned in the word family of the simplex words they contain to indicate the lexicographical relation between ground words and derivation. This means that with editing a dictionary entry of the ANW, the lexicographer simultaneously provides information of the words related to this entry.

2.4.2 Word Sketches and Sketch Grammar

Word sketches are one-page, automatic, corpus-based summaries of a word's grammatical and collocational behaviour, which are generated by the Sketch Engine. They improve on standard collocation lists by finding collocates in specific grammatical relations, and then producing one list of subjects, one of objects, etc. rather than a single grammatically blind list. In order to identify a word's grammatical and collocational behaviour, a Sketch Grammar is needed, which is basically a regular expression grammar defining patterns over POS-tags. The Dutch sketch grammar is completely geared to the ANW and models exactly those relations that lexicographers need to provide when editing an entry. The adaption of the Sketch Engine to Dutch including an overview of the Sketch Grammar has been described in Tiberius/Kilgariff (2009).

2.4.3 Automatic data acquisition

For the ANW, data on spelling, inflection, and hyphenation have been automatically inserted from the official Word list of the Dutch Language (*Woordenlijst der Nederlandse Taal*). We are currently exploring other categories for which data could be automatically acquired. Theoretically, it is also possible to provide automatically generated information on regional variation (Dutch in the Netherlands versus Dutch in Flanders), synonymy, sense distribution, and neologisms, but this automatically generated data always needs to be checked by the lexicographers. This makes automatic data acquisition not only part of the phase of data processing, but gives it also a place in the next phase, i.e., data analysis.

2.5 The phase of data analysis

In the analysis phase of the ANW project, lexicographers and lexicographic assistants work closely together. The lexicographic assistants prepare the dictionary entries before they go to the lexicographers. They also check the automatically compiled information (e.g. spelling) and they add (preliminary) information to a selected subset of the information categories that are present in the article structure, i.e., grammatical information and word family. The entries are then passed on to the lexicographers who turn them into complete dictionary articles by adding semagrams, definitions, combinations, idioms and proverbs, example sentences, etc., all based on an analysis of the corpus data in the Sketch Engine (concordances, word sketches, etc.). While this basically does not differ from writing entries for paper dictionaries, lexicographers working on online dictionaries have to do more, e.g. cross-references may not only be placed within one entry, but also between different entries in the dictionary, which means that the lexicographer may be editing two dictionary articles simultaneously.

Once the lexicographers have finished the articles, they are passed back to the lexicographic assistants who do a final check of the complete article, check the example sentences and add multimedia (i.e., audio and/or video files) based on specifications from the lexicographers. Klosa (2013, p. 520) already includes the task of adding multimedia in the phase of processing. In the ANW it is still mainly a manual task, so in our view it should be included under the phase of analysis. The articles are then ready for proofreading which is done by the editor in chief and the project manager.

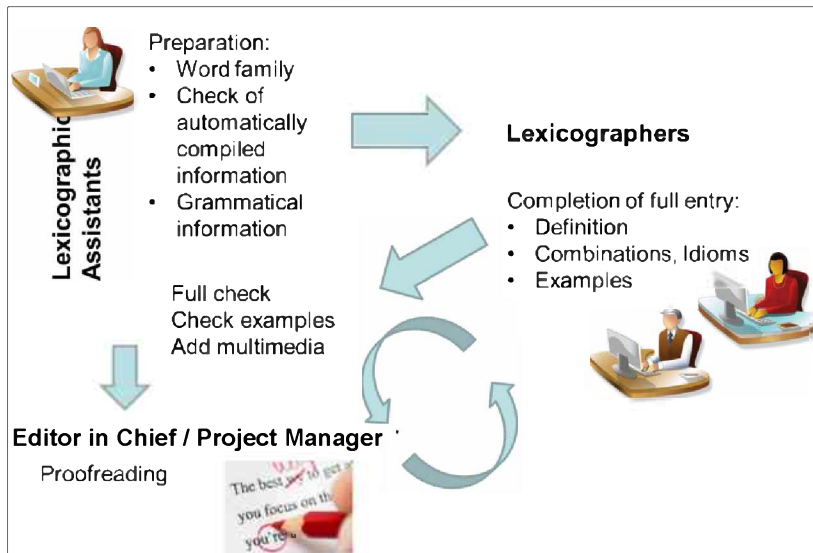


Figure 3: The ANW work flow

After proofreading the articles go back to the lexicographers to carry out the final corrections. Once that is done, the editor in chief and the project manager do a final check after which they change the status of the dictionary articles to ‘ready for online’ in the database. Every three months a new online version of the dictionary is created. Automatic checks are carried out for spelling and broken links which are then corrected manually. After that, the new version of the dictionary is put on an application test environment where it stays for one week. During that period, errors, inconsistencies, poor definitions, etc. can still be corrected. If after one week, the version on the application test environment is approved, an update of the dictionary is released on the public web site. For the first release in December 2009, the dictionaries’ accompanying texts also had to be written.

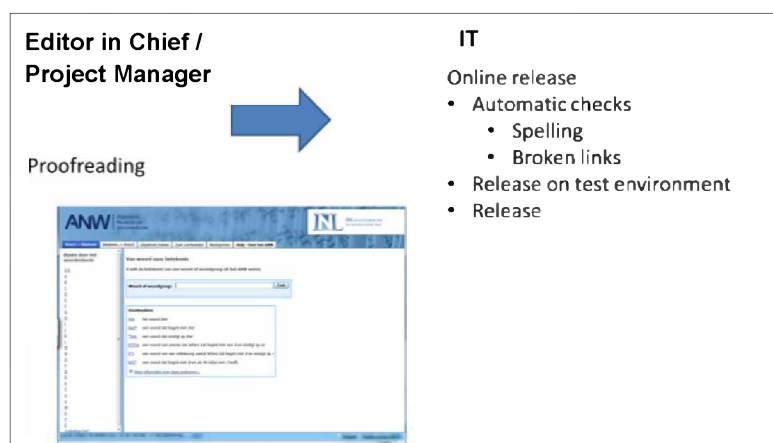


Figure 4: Phase of preparation for online release

3. Conclusion

In this paper we have sketched the lexicographical process of the ANW dictionary using the six phases distinguished by Klosa, i.e., preparation, data acquisition, computerisation, data processing, data analysis, and preparation for online release. We have shown that the preparation of an online dictionary is more complex than that of a paper dictionary. The different phases generally follow nicely onto each other for a paper dictionary, whereas this is not the case in a computer lexicographical process. Here a certain reiteration through the different phases can often be seen. For the ANW, at the moment this applies especially to the last two phases – the phase of analysis and the phase of preparation for online release. However, if we move from a static to a dynamic corpus as a source of our dictionary as we are currently setting up, more phases will be involved in the reiteration. We will then get a cycle of phase two (data acquisition) to phase six (preparation for online release). The complete process is illustrated in Table 1 for the ANW. It clearly shows that it is necessary to rethink the lexicographical process with an online product in mind in order to realise an efficient workflow and a smooth lexicographical process, as many factors are involved.

	1	5	10	15
Preparation				
Data Acquisition				
Computerisation				
Data Processing				
Data Analysis				
Preparation Online Release				

Table 1: Phases in the computer lexicographical process for the ANW

References

- Beijk, Egbert/Colman, Lutgarde/Moerdijk, A. F. M. J. (Fons) (eds.) (2009): *Fons verborum. Feestbundel voor prof. dr. A.F.M.J. (Fons) Moerdijk, aangeboden door vrienden en collega's bij zijn afscheid van het Instituut voor Nederlandse Lexicologie*. Leiden/Amsterdam: Instituut voor Nederlandse Lexicologie/Gopher BV.
- Bernal, Elisenda/De Cesaris, Janet (eds.) (2008): *Proceedings of the XIII Euralex International Congress, Barcelona, Universitat Pompeu Fabra, 15.-19. Juli 2008*. (= Sèrie activitats 20). Barcelona: Documenta Universitaria.
- De Does, Jesse/Van der Voort van der Kleij, John (2002): *Tagging the Dutch PAROLE corpus*. In: Theune, Mariet/Nijholt, Anton/Hondorp, Hendri (eds.): *Computational linguistics in the Netherlands 2001. Selected Papers from the Twelfth CLIN Meeting*. Amsterdam/New York: Rodopi, pp. 62-76.
- Hanks, Patrick (2002): *Mapping meaning onto use*. In: Corréard, Marie-Hélène (ed.): *Lexicography and natural language processing: A festschrift in honour of B.T.S. Atkins*. Stuttgart: Euralex, pp. 156-198.
- Kilgariff, Adam/Rychly, Pavel/Smrz, Pavel/Tugwell, David (2004): *The sketch engine*. In: Williams, Geoffrey/Vessier, Sandra (eds.): *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6-10, 2004*. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud, pp. 105-116.
- Kilgariff, Adam/Husák, Milos/McAdam, Katy/Rundell, Michael/Rychlý, Pavel (2008): *GDEX: Automatically finding good dictionary examples in a corpus*. In: Berndal/De Cesaris (eds.), pp. 425-432.
- Kilgariff, Adam/Kovar, Vojtech/Rychlý, Pavel (2010): *Tickbox Lexicography*. In: Granger, Sylviane/Paquot, Magali (Hg.): *eLexicography in the 21st century: New challenges, new applications*, Louvain-la-Neuve, 22-24 October 2009, Centre for English Corpus Linguistics, Université catholique de Louvain. Louvain-la-Neuve: CECL, pp. 115-118.

- Klosa, Annette (2013): The lexicographical process (with special focus on online dictionaries). In: Gouws, Rufus Hjalmar/Heid, Ulrich/Schweickard, Wolfgang/Wiegand, Herbert Ernst (eds.): *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography.* (= *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 5.4*). Berlin/Boston, pp. 517-524.
- Landau, Sidney (1984): *Dictionaries: The art and craft of lexicography*. New York: The Scribner Press.
- Moerdijk, Fons (2008): Frames and Semagrams. Meaning Description in the General Dutch Dictionary. In: Berndal/De Cesaris (eds.), pp. 561-571.
- Niestadt, Jan (2009): De ANW-artikeleditor: software als strategie. In: Beijck et al. (eds.), pp. 215-222.
- Referentiebestand Belgisch-Nederlands (2005): Den Haag: Nederlandse Taalunie.
- Referentiebestand Nederlands (2005): Den Haag: Nederlandse Taalunie.
- Schröder, Martin (1997): Brauchen wir ein neues Wörterbuchkartell? Zu den Perspektiven einer computerunterstützten Dialektlexikographie und eines Projektes „Deutsches Dialektwörterbuch“. In: *Zeitschrift für Dialektologie und Linguistik* 64, 1, pp. 57-66.
- Schoonheim, Tanneke/Tempelaars, Rob (2010): Dutch lexicography in progress: the Algemeen Nederlands Woordenboek. In: Dykstra, Anne/Schoonheim, Tanneke (eds.): *Proceedings of the XIV Euralex International Congress* (Leeuwarden, 6-10 July 2010). Leeuwarden: Afûk, pp. 718-725.
- Tiberius, Carole/Kilgariff, Adam (2009): The sketch engine for Dutch with the ANW corpus. In: Beijck et al. (eds.), pp. 237-255.
- Van Dale (2010): *Groot Woordenboek van de Nederlandse Taal*. 14th edition. Eds. Ton den Boon, Nicoline van der Sijs, Dirk Geeraerts. Utrecht: Van Dale Lexicografie.
- Wiegand, Herbert Ernst (1999): *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilbd. Berlin/New York: de Gruyter.
- Woordenlijst der Nederlandse taal (2005): Antwerpen/Den Haag: Lannoo Uitgeverij/SDU Uitgevers.

Annette Klosa

Der lexikographische Prozess im Projekt *elexiko*

1. Einführung

elexiko,¹ ein Onlinewörterbuch zur deutschen Gegenwartssprache, wurde von Anfang an für die Publikation im Internet geplant und realisiert. Es ist also im wahrsten Sinne des Wortes ein Internetwörterbuch, für das außerdem charakteristisch ist, dass es noch nicht vollständig vorliegt, dennoch aber schon publiziert wird. Man kann bei *elexiko* daher von einem „Ausbauwörterbuch“ (Schröder 1997, S. 60) oder einem „dynamischen Wörterbuch“ (Lemberg 2001, S. 81) sprechen, nach dem Kriterium der Vollständigkeit würden Storrer/Freese (1996) sowie Storrer (1998 und 2001) es als „Wörterbuch im Aufbau“ einordnen. Dieses Merkmal hat einen großen Einfluss auf den lexikographischen Prozess von *elexiko*, indem bei solchen Wörterbüchern wie *elexiko* dieser im Grunde nicht abgeschlossen ist bzw. zu Ende geführt werden kann, sondern so lange andauert, wie das Wörterbuch publiziert wird.

Da *elexiko* sowohl durchgehend mit Unterstützung von Computern erarbeitet wird als auch auf/in Computern publiziert wird, ist der lexikographische Prozess in diesem Fall genauer als computerlexikographischer Prozess (nach Wiegand 1998, S. 233ff.) zu bestimmen. In Fortführung von Wiegands Vorschlägen werden für computerlexikographische Prozesse nach Klosa (2013) die Phasen der Vorbereitung und Planung, der Datenbeschaffung, der Computerrisierung, der Datenaufbereitung, der Datenauswertung und der Vorbereitung für den Onlinerelease unterschieden. Charakteristisch für Wörterbücher wie *elexiko* ist dabei, dass alle Phasen nicht notwendigerweise hintereinander stattfinden, sondern dass sie auch parallel zueinander ablaufen können. Eine klare Grenzziehung zwischen den einzelnen Phasen ist daher nicht immer möglich.

Ein zweites Charakteristikum für solche Wörterbücher im Aufbau ist, dass in der Phase der Datenauswertung nicht von A bis Z vorgegangen werden muss, sondern dass die Stichwörter des Wörterbuches nach anderen Kriterien als ihrer alphabetischen Abfolge für die Bearbeitung und sukzessive Publikation ausgewählt werden können. Die Erarbeitung des Wörterbuches erfolgt dann in sogenannten Modulen (Haß 2005, S. 13ff.; Klein 2004, S. 300f.). Verschiedene Module eines Onlinewörterbuches können sich in verschiedenen Phasen des lexikographischen Prozesses befinden, in dem für ein Modul beispielsweise noch Daten beschafft werden müssen, während die Stichwörter eines anderen Moduls schon online erscheinen. Und sogar innerhalb eines Moduls können sich verschiedene Stichwörter in verschiedenen Phasen der lexikographischen Prozesse befinden, z.B. im folgenden Fall, der auch für *elexiko* zutrifft:

Modules not only as a whole but also single entries within a module can be situated in different phases if the dictionary allows online release of single entries. Thus, a number of entries in the same module may already be online while still others are in the phase of data analysis. (Klosa 2013, S. 522).

¹ Vgl. www.elexiko.de (zuletzt eingesehen: 5.2.2015). Zur Konzeption von *elexiko* vgl. Haß (Hg.) (2005). Zur praktischen Umsetzung dieser Konzeption vgl. Klosa (Hg.) (2011).

2. Die Phasen des computerlexikographischen Prozesses von *elexiko*

2.1 Vorbereitungsphase

In der Vorbereitungsphase für *elexiko*, dessen Projektlaufzeit im Jahr 1997 begonnen hat, wurde viel Zeit in die Entwicklung der inhaltlichen Konzeption und in etliche Pilotstudien investiert. Die Ergebnisse dieser Bemühungen sind im Band „Grundfragen der elektronischen Lexikografie. *elexiko* – Das online-Informationssystem zum deutschen Wortschatz“ (Haß (Hg.) 2005) publiziert, in dem sowohl die Makrostruktur des Wörterbuches wie auch die Mikrostruktur ausführlich dargelegt werden. Fragen der Datenmodellierung, der Korpuszusammenstellung sowie der Mediostruktur werden ebenso erläutert wie erste Überlegungen für die Onlineoberfläche u.Ä. Erst später wurde allerdings mit der Umsetzung der inhaltlichen Konzeption in ein lexikographisches Redaktionshandbuch begonnen (vgl. Klosa 2011a, S. 15).

Das Konzept für die Wörterbuchinhalte wurde erprobt, indem zahlreiche Musterartikel (im sogenannten *elexiko*-Demonstrationswortschatz) verfasst wurden. Dieser Demonstrationswortschatz enthält knapp 250 Stichwörter, bei deren Auswahl zwei Kriterien zur Anwendung kamen: Ausgewählt wurden erstens die mithilfe des Verfahrens der Kookkurrenzanalyse² ermittelten statistisch signifikanten Kookkurrenzpartner, also Kollokatoren des Nomens *Mobilität*³. Das zentrale Wort *Mobilität* wurde gewählt, weil die Diskussion hierüber, vor allem im beruflichen Kontext, im *elexiko*-Korpus einen breiten Raum einnimmt und auch allgemein wichtig und aktuell ist, ohne zu sehr religiös oder weltanschaulich gefärbt zu sein. Zweitens wurde diese Menge systematisch ergänzt nach den aus anderen bzw. älteren Wörterbüchern bekannten Proportionen von Wortarten, Alphabetstrecken und diversen Wortbildungstypen. Durch diese Auswahl bzw. ihre systematische Ergänzung war es möglich, Schätzungen für die Bearbeitungsdauer des später begonnenen Moduls „Lexikon zum öffentlichen Sprachgebrauch“⁴ anzufertigen. Die Erarbeitung des Demonstrationswortschatzes gehört inhaltlich zwar noch in die Vorbereitungsphase, weil im Grunde erst nach Abschluss dieser Arbeiten am systematischen weiteren Ausbau von *elexiko* gearbeitet werden konnte bzw. wurde, greift natürlich aber auch schon auf die später beschriebenen Phasen im lexikographischen Prozess aus.

Die Arbeit am Projekt *elexiko* wurde zunächst ohne vollständige Durchdringung computerlexikographischer Prozesse, dafür mit viel Pioniergeist und opportunistischer Ausnutzung der am IDS gegebenen Möglichkeiten durch die beteiligten lexikographischen Mitarbeiter(innen) aufgenommen. Dies lässt sich u. a. daran ablesen, wie die Projektschritte von 1997 bis 2004 in Haß (2005, S. 13ff.) beschrieben werden:

Erste Skizzen zu einem lexikalischen Informationssystem entstanden Ende 1997. Der Aufbau einer Arbeitsgruppe, die Exploration der computertechnischen Möglichkeiten und die Konkretisierung eines den personellen wie technischen Möglichkeiten angepassten Konzepts mündeten in die linguistische Entwicklung und texttechnologische Realisierung einer Wortartikelstruktur in Form einer ungewöhnlich komplexen DTD-Struktur. [...]

In der darauffolgenden Phase wurde diese DTD getestet, evaluiert und insgesamt dreimal überarbeitet. [...] In diese Konzeptionsphase fallen sämtliche Bemühungen um eine mittelfristig stabile Lösung der gesamten Softwarearchitektur eines solchen [sic!] Vorhabens, die auch heute nicht und vermutlich nie ganz wunschgemäß gegeben ist. Anfang 2003 war die Konzeptionsphase sowohl linguistisch-lexikografisch als auch computertechnisch so weit abgeschlossen, dass mit der Realisierung erster Wortschatzbereiche (Module [...]), d.h. mit dem Verfassen von Wortartikeln und ihrer Speicherung in der Datenbank begonnen werden konnte.

² Zur Methode der Kookkurrenzanalyse vgl. www.ids-mannheim.de/kl/projekte/methoden.html (zuletzt eingesehen: 5.2.2015).

³ Vgl. das Stichwort *Mobilität* online unter www.owid.de/artikel/62377 (zuletzt eingesehen: 5.2.2015).

⁴ Vgl. www.owid.de/wb/elexiko/projekt/modSpracheb.html (zuletzt eingesehen: 5.2.2015).

In die Realisierungsphase ist auch die Erarbeitung der knapp 300.000 Wörter umfassenden Stichwortliste aus dem IDS-Korpus einbezogen, [...].

Seit Januar 2004 sind diese 300.000 gebräuchlichen Einwortausdrücke mit orthografischen und morphologischen Angaben öffentlich zugänglich. [...]

ellexiko wird auch in Zukunft modular weiterentwickelt und ausgebaut, d. h. nicht entlang dem Alphabet, sondern vertikal nach Teilwortschätzen bzw. horizontal nach bestimmten Angabearten. [...] Das erste Beispiel für einen vertikal ausgearbeiteten Teilwortschatz ist der 240 Lexeme umfassende Demonstrationswortschatz. [...]

Seit Juni 2004 stehen die 240 Artikel des sog. Demonstrationswortschatzes online zur Verfügung; [...]. (Haß 2005, S. 13ff.)

Zu Ende der 90er Jahre des vorigen Jahrhunderts wurde mit dem geplanten Wörterbuch im Bereich der wissenschaftlichen Lexikographie Neuland betreten. Es gab keine Vorbilder für vollständig neu zu erarbeitende, kontinuierlich zu erweiternde Onlinewörterbücher, sodass es im Grunde keine Erfahrungswerte gab, auf denen das Projekt hätte aufbauen können. Allerdings lagen damals durchaus schon Veröffentlichungen vor, die sich mit dem lexikographischen Prozess an sich befassen (z.B. Dubois 1990; Landau 1984), und es gab Ansätze zur Analyse der Einsatzmöglichkeiten des Computers in der Lexikographie (z.B. Knowles 1990). Trotzdem wurde im Projekt *ellexiko* in der Vorbereitungsphase auch einiges versäumt, vor allem die Erstellung eines Organisationsplans mit Angaben zu Finanzierung, Personal (auch technischem Personal!), Zeitplan und Workflow, der als zweiter wichtiger Bestandteil des Wörterbuchplans das inhaltliche Konzept für *ellexiko* hätte ergänzen müssen. Ebenfalls versäumt wurde eine gründliche Marktanalyse der Ende der 1990er Jahre existierenden lexikographischen Redaktionssysteme. Beide Versäumnisse haben in Folge die Erarbeitung von *ellexiko* zeitlich und finanziell belastet, indem es zu eigentlich unnötigen Investitionen von Zeit und Geld kam, um eingangs schlecht oder gar nicht geplante Umstände zu korrigieren (z.B. kein Rückgriff auf ein käufliches Redaktionssystem, sodass später erhebliche Personalmittel und Zeit aufgewandt werden mussten, um eine eigene Redaktionsumgebung zu schaffen (vgl. Abel/Klosa 2012, S. 417f.)). Schließlich wurde die Planung des Onlinedesigns des Wörterbuches ohne Unterstützung von Webdesignspezialisten durchgeführt und verschiedene Benutzungsmöglichkeiten von *ellexiko* (z.B. erweiterte Suche, Angebot der reinen Stichwortliste in links- und rechtsalphabetischer Sortierung) wurden ohne die Durchführung von Benutzungsstudien⁵ entworfen.

2.2 Datenbeschaffung

Die Phase der Datenbeschaffung war im Projekt *ellexiko* vor allem der Zusammenstellung des *ellexiko*-Korpus als virtuelles Korpus aus DEREKO, dem Deutschen Referenzkorpus des IDS Mannheim,⁶ gewidmet. Das Korpus besteht aus Texten unter anderem aus folgenden Zeitungen und Zeitschriften: Berliner Morgenpost, Berliner Zeitung, Der Spiegel, Die Presse, die tageszeitung, Die Welt, Die Zeit, Frankfurter Allgemeine Zeitung, Frankfurter Rundschau, Hamburger Morgenpost, Kleine Zeitung, Mannheimer Morgen, Neue Kronen-Zeitung, Oberösterreichische Nachrichten, Salzburger Nachrichten, St. Galler Tagblatt, Tiroler Tageszeitung, Vorarlberger Nachrichten und Züricher Tagesanzeiger. Es verbindet also bundesdeutsche, österreichische und Schweizer Quellen in einer den jeweiligen Sprecherzahlen in etwa entsprechenden Menge. Da es sich bei dem *ellexiko*-Korpus aber um ein dynamisches Korpus (bzw. ein sogenanntes Monitorkorpus) handelt, welches regelmäßig erweitert und ausgebaut

⁵ Einige Elemente der Onlineoberfläche von *ellexiko* wurden zwischenzeitlich in zwei Benutzungsstudien untersucht, vgl. hierzu Klosa/Koplenig/Töpel (2011).

⁶ Vgl. www.ids-mannheim.de/kl/projekte/korpora.html (zuletzt eingesehen: 5.2.2015).

wurde, ist das hier vorgestellte Wörterbuchprojekt immer wieder (zuletzt 2013) in die Phase der Datenbeschaffung zurückgekehrt, um neue Quellen zu ergänzen.

Parallel wurden sekundäre Quellen (andere gedruckte wie elektronische Wörterbücher) und tertiäre Quellen (z.B. Grammatiken)⁷ gesichtet und beschafft, die allerdings im Gesamtkontext des Projektes eine wesentlich unbedeutendere Rolle als das Korpus selbst einnehmen, weil die lexikographische Beschreibung in *ellexiko* generell einem korpusgestützten Ansatz verpflichtet ist (vgl. Klosa 2011a, S. 16f.). Auch diese Teilaufgabe in der Phase der Datenbeschaffung war nicht zu einem bestimmten Zeitpunkt abgeschlossen, sondern der inzwischen über 10-jährigen Projektlaufzeit ist geschuldet, dass immer wieder neu erschienene sekundäre und tertiäre Quellen ergänzt wurden. An diesem kleinen Beispiel zeigt sich (und dies gilt vermutlich für generell jeden lexikographischen Prozess), dass die theoretisch angenommenen Phasen nicht wirklich eine nach der anderen ablaufen, sondern sich immer wieder überlappen können.

In einigen Wortartikeln des *ellexiko*-Demonstrationswortschatzes waren schon Illustrationen enthalten; das entsprechende Bildmaterial wurde in der Phase der Datenbeschaffung besorgt. Seitdem hat sich allerdings gerade in diesem Bereich viel getan: Zwischenzeitlich wurden weitere Phasen der Datenbeschaffung (zuletzt 2010 bis 2013) eingeschoben, und zwar beispielsweise zur systematischen Gewinnung von Bildmaterial und zur Auswahl von Hördateien, welche das nach Abschluss des Demonstrationswortschatzes begonnene und vorangetriebene Modul „Lexikon zum öffentlichen Sprachgebrauch“ bereichern sollten. Zur Sammlung des Bildmaterials ist anzumerken:

Daneben wurde überprüft, welche Möglichkeiten es zur kostenlosen Gewinnung von Illustrationen gibt. Hierbei sind verschiedene Online-Datenbanken mit Zeichnungen, Fotografien, Comics und Videos ausgewertet worden (z.B. pixelio.de, Wikimedia Commons). Das Ergebnis ermutigt, was die Menge kostenlos zur Verfügung stehender Illustrationen (vor allem Fotos) betrifft, wirft aber auch einige neue Probleme auf: So sind die Fotos hinsichtlich ihrer Art, Qualität, Auflösung und Größe sehr unterschiedlich. Ebenso unterschiedlich sind die Vorgaben dazu, wie die Quellenangabe erfolgen muss. (Klosa 2011b, S. 164)

Die Hörbelege, die in *ellexiko* die natürlichsprachige Aussprache eines Stichwortes im Kontext demonstrieren, wurden zur gleichen Zeit aus der „Datenbank für gesprochenes Deutsch“ im „Archiv für gesprochenes Deutsch (2009)“⁸ des IDS ermittelt. Mit dem Jahr 2013 wurde die nachträglich eingeschobene Phase der Beschaffung von Bild- und Tondaten abgeschlossen.

2.3 Computerisierung

Zwei besonders zeitaufwändige Arbeitsschritte in der Phase der Computerisierung, nämlich die digitale Aufbereitung von Korpus-texten (Annotierung, Lemmatisierung) sowie die Programmierung eines Korpusrecherchetoools, konnten im Projekt *ellexiko* insofern entfallen, als das Projekt ein virtuelles Korpus aus dem „Deutschen Referenzkorpus (DEREKO)“ des IDS zusammenstellen konnte und für die Auswertung des Korpus auf COSMAS II,⁹ das Korpusrecherche- und -analysewerkzeug des IDS zurückgreifen kann. Im Projekt werden außerdem die in der Kookkurrenzdatenbank CCDB¹⁰ des IDS dokumentierten Ergebnisse berücksichtigt.¹¹

⁷ Zu primären, sekundären und tertiären Quellen in der Lexikographie vgl. Wiegand (1998, S. 140).

⁸ Vgl. <http://agd.ids-mannheim.de/datenbanken.shtml> (zuletzt eingesehen: 5.2.2015).

⁹ Vgl. <https://cosmas2.ids-mannheim.de/cosmas2-web/> (zuletzt eingesehen: 5.2.2015).

¹⁰ Vgl. <http://corpora.ids-mannheim.de/ccdb/> (zuletzt eingesehen: 5.2.2015) und Belica (2011).

¹¹ Vgl. hierzu genauer Storzjohann (2011).

Zur redaktionellen Arbeitsumgebung im Projekt gehören daneben die folgenden Komponenten (vgl. hierzu Abel/Klosa 2012, S. 416f.):

- ein XML-Editor (zunächst XMetal, jetzt Oxygen), in dem die lexikographischen Daten ediert werden,
- ein Verweismanager (*vernetziko*),¹² der die Erstellung und Pflege konsistenter Vernetzungen im Wörterbuch ermöglicht und außerdem die lexikographische Arbeit in anderer Hinsicht unterstützt (z.B. durch die Möglichkeit, Textbausteine anzulegen oder durch Abfragemöglichkeiten nach dem Bearbeitungsstand einzelner Einträge),
- eine Schnittstelle (EDAS – Electronic Dictionary Administration System)¹³ zur Datenbank (Oracle), in der die Daten gespeichert werden, in der sie durchsuchbar sind und aus der sie für die Online-Präsentation exportiert werden,
- über EDAS außerdem Zuweisung eines bestimmten Bearbeitungsstatus zu Wortartikeln und Durchführung komplexer Suchanfragen im Wortartikelbestand,
- interne Online-Ansicht der Wörterbuchoberfläche innerhalb des Portals OWID zur Überprüfung vor der Freischaltung,
- ein elektronisches Redaktionshandbuch und eine DTD-Dokumentation.

Im Projekt *ellexiko* wird also kein lexikographisches Redaktionssystem im eigentlichen Sinn eingesetzt, doch funktioniert die Arbeitsumgebung nach einiger Zeit der Entwicklung und zwischenzeitlichen Optimierungen (hier ist insbesondere das Vernetzungstool zu nennen) trotzdem zufriedenstellend.

2.4 Datenaufbereitung

Ein großer Arbeitsbereich in der Phase der Datenaufbereitung war die Erstellung der Stichwortkandidatenliste aus dem *ellexiko*-Korpus. Dabei wurde folgendermaßen vorgegangen:¹⁴

1. Mithilfe eines automatischen Lemmatisierers wurden die Flexions- bzw. Paradigmenformen von Wörtern aus den Texten der IDS-Korpora geschriebener Gegenwartssprache Wortformen zugewiesen, die gemäß der Annahmen des automatischen Lemmatisierers als 'wörterbuchübliche Grund- oder Nennformen' gedeutet werden können, z.B. dem Infinitiv bei Verben, dem Nominativ Singular bei Nomen, dem Positiv bei Adjektiven usw.; das sind die sog. Stichwortkandidaten. Zusätzlich wurden auch lexikografisch relevante Flexionsformen eigens erfasst, beispielsweise Partizipien, unregelmäßige Pluralbildungen, die erste Person Präteritum der starken Verben u.Ä. [...].
2. Um von vornherein die Zahl insbesondere formal fehlerhaft angesetzter Stichwortkandidaten einzugrenzen, erfolgte ein sichernder Korrekturabgleich der Stichwortkandidatenliste mit Listen anderer Wörterbücher.

Die Stichwortkandidatenliste ist das Ergebnis eines Zusammenspiels der beiden Methoden 'Korpusextraktion mit automatischer (Grund-) Formzusammenführung' und 'korrektiver formaler Abgleich mit vorhandenen Wörterbuchlisten' im Anschluss daran. Das Resultat wäre allerdings noch um ein Vielfaches umfangreicher, hätte man auf dieser Stufe nicht eine dritte, quantitativ einschränkende, also auswählende Größe berücksichtigt: die Frequenz.

Die weitgehend automatisch generierte Liste enthält deshalb 'nur' knapp 320.000 Stichwortkandidaten, die mit einer Mindestfrequenz von acht im Korpus vorkommen und in anderen Wörterbüchern enthalten sind. Die Frequenz von acht bedeutet, dass das betreffende Wort in seiner Grund- bzw. Nennform und/oder in den dazugehörigen Flexionsparadigmenformen mindestens achtmal im Korpus belegt ist. Die Zahl acht ist dabei zufällig jene Größe, mit deren Hilfe ein Schnitt bei der Extraktion aus dem Korpus vorgenommen werden konnte, durch den sich der Umfang der Stichwortkandidatenliste in idealer Weise beschränken ließ: [...]. (vgl. Schnörch 2005, S. 75f.)

¹² Vgl. Meyer/Müller-Spitzer (2010) und Meyer (2011).

¹³ Entwickelt im früheren Projekt Texttechnologie, jetzt Grammis II: Grammatische Datenbanken und Informationssysteme, vgl. www.ids-mannheim.de/gra/projekte/grammis2.html (zuletzt eingesehen: 5.2.2015) und Müller-Spitzer/Schneider (2009).

¹⁴ Verantwortlich für diesen Arbeitsschritt war der Programmbereich Korpuslinguistik (damals noch „AG Korpustechnologie“) des IDS, siehe www.ids-mannheim.de/kl/projekte.html (zuletzt eingesehen: 5.2.2015).

Ein wichtiger weiterer Schritt war die (automatische) Ermittlung von Frequenzen der Stichwörter, und zwar die Summe aller Vorkommen der jeweiligen Grundform und die Summe aller dieser Grundform zugeordneten Flexionsformen. Zwar können diese Frequenzangaben „bei einem dynamischen Monitorkorpus wie dem *ellexiko*-Korpus immer nur eine Momentaufnahme sein“ (Klosa 2011b, S. 158), doch wurden sie genutzt, um sogenannte Frequenzschichten zu definieren, die wiederum die Grundlage für die Definition von zu bearbeitenden Stichwortmengen (Modulen) in *ellexiko* bilden (z.B. sind im „Lexikon zum öffentlichen Sprachgebrauch“ Stichwörter mit einer Frequenz zwischen 10.000 und 500.000 im Korpus enthalten).

Schließlich gehört in diese Phase des lexikographischen Prozesses auch die Umsetzung der Modellierung der lexikographischen Daten in eine entsprechende Datenbankstruktur (vgl. Wiegand 1998, S. 238). Testweise wurde hierbei zunächst mit der Datenbank Tamino,¹⁵ später dann mit Oracle¹⁶ gearbeitet. Bei der Entwicklung des XML-basierten Datenbanksystems konnte das Projekt von den für grammis (das grammatische Informationssystem des IDS) gewonnenen Erfahrungen des Projektes „Texttechnologie und Datenbanken“ profitieren (vgl. Müller-Spitzer/Schneider 2009).

2.5 Datenauswertung

In der Phase der Datenauswertung können sich automatisch-korpuslinguistische sowie redaktionell-lexikographische Arbeiten verbinden. Im Projekt *ellexiko* wurden z.B. Belege aus dem *ellexiko*-Korpus ermittelt, die bei all jenen Stichwörtern in *ellexiko* angezeigt werden, die noch nicht redaktionell bearbeitet sind.¹⁷ Hierbei, wie auch bei weiteren automatisch erzeugten lexikographischen Angaben, wird aber auf völlig automatische Integration der analysierten Daten in die Wortartikel verzichtet, sondern wo immer möglich wird ein redaktioneller Prüfung (wenigstens stichprobenartig) durchgeführt und entsprechende Korrekturen werden vorgenommen.¹⁸

In die Phase der Datenauswertung fallen insbesondere redaktionelle Arbeitsgänge. Hierzu zählte im Projekt *ellexiko* die redaktionelle Prüfung der automatisch erstellten Stichwortkandidatenliste (s.o.).¹⁹ Hauptaufgabe in dieser Phase eines jeden lexikographischen Prozesses ist das Verfassen neuer Wortartikel, und dieser Arbeitsschritt umfasst auch im Projekt *ellexiko* den längsten Zeitraum in der Projektlaufzeit (nämlich kontinuierlich seit 2003).²⁰ Dabei werden die Korpusdaten (in Form von Kookkurrenzlisten als Ergebnis der Kookkurrenzanalyse,²¹ KWIC-Zeilen, Frequenzlisten usw.) analysiert und interpretiert, die Lesarten werden disambiguiert, Definitionen geschrieben, Korpusbelege zur Integration in den Wortartikel ausgesucht, Sekundärquellen geprüft usw. Diese Arbeiten laufen dabei im Grundsatz nicht anders ab, als sie es für ein Printwörterbuch tun würden:

¹⁵ Vgl. www.softwareag.com/corporate/products/az/webmethods/default.asp (zuletzt eingesehen: 5.2.2015).

¹⁶ Vgl. www.oracle.com/de/products/database/overview/index.html (zuletzt eingesehen: 5.2.2015).

¹⁷ Vgl. hierzu genauer Klosa (2011a, S. 18ff.).

¹⁸ Zur Verbindung redaktionell erarbeiteter wie automatisch ermittelter Angaben in Wörterbüchern vgl. genauer Klosa (2010).

¹⁹ Vgl. genauer Schnörch (2005, S. 77f.).

²⁰ Einen genauen Einblick in die redaktionelle Arbeit im Projekt *ellexiko* geben die in Klosa (Hg.) (2011) enthaltenen Beiträge.

²¹ Vgl. www.ids-mannheim.de/kl/projekte/methoden/ka.html (zuletzt eingesehen: 5.2.2015).

Alle genuin lexikographischen, philologischen und linguistischen Kenntnisse und Fertigkeiten, welche nötig sind, um ohne Computereinsatz in der Phase der Materialauswertung Handlungen vom Typ EINEN WÖRTERBUCHARTIKEL SCHREIBEN erfolgreich ausführen zu können, werden auch benötigt, um in der Phase der Datenauswertung Handlungen vollziehen zu können, die zum Typ EINEN WÖRTERBUCHARTIKEL COMPUTERUNTERSTÜTZT SCHREIBEN gehören. (Wiegand 1998, S. 239)

Der Unterschied bei der Erarbeitung eines Internetwörterbuches zeigt sich eher dann, wenn für ein Wörterbuch wie *ellexiko* nicht nur (elektronische) Vernetzungen zwischen einzelnen Wortartikeln oder den Wortartikeln mit den lexikographischem Umtexten angelegt werden, sondern auch zu externen Quellen (z.B. Enzyklopädien) verlinkt wird.

In die Phase der Datenauswertung fällt auch ein Arbeitsgang, in dem die während der Datenbeschaffungsphase zusammengetragenen multimedialen Elemente (Illustrationen, Hördateien) gesichtet bzw. angehört und getestet werden, damit die für die Wortartikel geeigneten ausgewählt werden können. Bild- und Tondaten werden außerdem analysiert, um die für die Wortartikel gewünschten Bild- und Tonausschnitte festlegen zu können.

2.6 Vorbereitung für Onlinerelease

Vor der Freischaltung von Wörterbüchern im Internet sind eine Menge technischer wie inhaltlicher Arbeitsschritte zu leisten, damit das Wörterbuch in möglichst überzeugender Qualität genutzt werden kann. So ist natürlich nötig, dass die Präsentation der Wortartikel im Internet technisch durch entsprechende Programmierung ermöglicht wird. Im Projekt *ellexiko* wird dies mithilfe von Stylesheets, welche die Umwandlung der XML-Inhalte für die Ansicht in HTML steuern, realisiert:

Der redaktionelle Input bestand hier im Entwurf der Wortartikelansichten und der Festlegungen dazu, welche Angaben online wie erscheinen sollen, sowie im gründlichen Testen der Umsetzung dieser Vorgaben. (Klosa 2011a, S. 22)

Umfänglich müssen auch alle Wortartikel vor ihrer Freischaltung im Internet getestet sowie inhaltlich und formal Korrektur gelesen werden. In *ellexiko* werden alle Wortartikel zunächst einer doppelten inhaltlichen Korrektur sowie anschließend einer formalen Korrektur unterzogen. Alle vorgenommenen Korrekturen werden nochmals kontrolliert, und vor dem Erscheinen des fertigen Artikels im Internet werden generell alle Hyperlinks im Wortartikel überprüft, Illustrationen werden testweise geöffnet, Tondateien testweise abgespielt usw.

Während dies bei einem Ausbauwörterbuch wie *ellexiko* ein fortwährender Prozess ist, war das Verfassen der Benutzungshinweise, des Glossars und weiterer Wörterbuchaußentexte ein geschlossenes Arbeitspaket, das vor der Veröffentlichung der ersten Wortartikel des „Lexikons zum öffentlichen Sprachgebrauch“ eingeschoben wurde. Zwischenzeitlich (2007) sind diese Texte allerdings schon wieder gründlich überarbeitet und ergänzt worden.

3. Kritische Betrachtung des lexikographischen Prozesses in *ellexiko*

Wie eingangs schon erwähnt wurde, wurde mit der Arbeit am Projekt *ellexiko* begonnen, ohne den für die Realisierung notwendigen computerlexikographischen Prozess vollständig zu durchdenken, sodass es zu Versäumnissen hinsichtlich wichtiger planerischer Schritte kam. Zugleich ist aber festzuhalten, dass erst in der praktischen Arbeit an diesem vollständig neu aus Korpusdaten erarbeiteten, ausschließlich für das Medium Internet konzipierten und erst-

mals als Wörterbuch im Aufbau publizierten Nachschlagewerk die nötigen Erfahrungen gesammelt werden konnten, um schließlich den lexikographischen Prozess in seiner Gänze erfahren und beschreiben zu können.

Da die Arbeit am Projekt *elexiko* nicht isoliert, sondern in einem ganz bestimmten Umfeld am IDS stattgefunden hat bzw. stattfindet, muss sich die Ausgestaltung des lexikographischen Prozesses unter Umständen an die hier vorhandenen Gegebenheiten anpassen. Dies hatte zum einen Auswirkungen auf die Inhalte (die grammatischen Angaben in *elexiko* sind beispielsweise konform mit den Angaben in *grammis*), zum anderen auf die eingesetzte Technik (z.B. auf die letztendliche Entscheidung für die Datenbank Oracle, die bereits für *grammis* eingesetzt wurde).

Generell ist der lexikographische Prozess von *elexiko* weniger dadurch zu beschreiben, dass seine Phasen hintereinander ablaufen, sondern eher dadurch, dass stets bestimmte Phasen des Prozesses parallel ablaufen und sich bestimmte Phasen auch wiederholen können. Der Prozess lässt sich im Grunde weniger linear, als in gewisser Weise zirkulär beschreiben. 2014 befand sich etwa ein Bearbeitungsteilwortschatz zu Wortbildungsmitteln noch in der Planungsphase, während sich das Modul „Lexikon zum öffentlichen Sprachgebrauch“ bereits in der Phase der Datenauswertung befand bzw. teilweise bereits veröffentlicht war. Auch einzelne Stichwörter in *elexiko* können sich in unterschiedlichen Phasen des Prozesses befinden: Ein und dasselbe Stichwort kann z.B. bereits mit automatisch generierten lexikographischen Angaben publiziert sein, für die nachgelagerte redaktionelle Bearbeitung setzt aber erst die Phase der Datenauswertung ein. Ebenso können sich Mengen von Stichwörtern in einem Bearbeitungsteilwortschatz in unterschiedlichen Phasen befinden. Derzeit läuft für alle redaktionell bearbeiteten und online freigeschalteten Stichwörter, die illustrierbar sind, teilweise noch die Phase der Datenbeschaffung, indem gezielt nach neuen oder weiteren Illustrationen gesucht wird. In dem Moment, in dem die Stichwörter nachträglich mit Illustrationen versehen werden, durchlaufen sie wieder die Phase der Datenauswertung (Prüfung der Abbildungen, Auswahl des Bildausschnittes, Erfassen der nötigen Informationen im Wortartikel) und der Vorbereitung für den Onlinerelease (Kontrolle der Darstellung, Überprüfung des Links zur Bildquelle).

Generell lässt sich aus der beschriebenen Komplexität des lexikographischen Prozesses im Projekt *elexiko* Folgendes schließen, was sicherlich auch auf andere Wörterbuchprojekte übertragen werden kann:

An online dictionary under construction is an open system (see Schröder 1997: 16). This also implies that organizing (calculating, segmenting, regulating) and controlling this process (see Wiegand 1998: 134) is of the utmost importance. Training the staff to work simultaneously on many different things in different stages is also helpful (as discussed by Landau 1984: 264). It is also very important to render this process transparent for users [...]. (Klosa 2013, S. 506).

4. Literaturverzeichnis

4.1 Fachliteratur

- Abel, Andrea/Klosa, Annette (2012): Der lexikographische Arbeitsplatz – Theorie und Praxis. In: Vatvedt Fjeld, Ruth/Torjusen, Julie Matilde (Hg.): Proceedings of the 15th EURALEX International Congress in Oslo 2012, S. 413-421.
- Belica, Cyril (2011): Semantische Nähe als Ähnlichkeit von Kookkurrenzprofilen. In: Abel, Andrea/Zanin, Renata (Hg.): Korpora in Lehre und Forschung. Bozen-Bolzano: Freie Universität, S. 155-178.
- Dubois, Claude (1990): Considérations générales sur l'organisation du travail lexicographique. In: Hausmann et al. (Hg.), S. 1574-1588.
- Haß, Ulrike (Hg.) (2005): Grundfragen der elektronischen Lexikographie. *elexiko* – das Online-Informationssystem zum deutschen Wortschatz. (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York: de Gruyter.
- Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (Hg.) (1990): Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie. 2. Teilbd. (= Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 5.2). Berlin/New York: de Gruyter
- Klein, Wolfgang (2004): Das digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts. In: Scharnhorst, Jürgen (Hg.): Sprachkultur und Lexikographie. (= Sprache – System und Tätigkeit 50). Frankfurt a.M. u.a.: Lang, S. 281-308.
- Klosa, Annette (2013): The lexicographical process (with special focus on online dictionaries). In: Gouws, Rufus/Hjalmar/Heid, Ulrich/Schweickard, Wolfgang/Wiegand, Herbert Ernst (Hg.): Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. (= Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 5.4). Berlin/Boston: de Gruyter, S. 517-524.
- Klosa, Annette (2011a): Einleitung. In: Klosa (Hg.), S. 9-26
- Klosa, Annette (2011b): Von Abbildung bis Wortelement: Weitere Ergänzungen und Änderungen in *elexiko*. In: Klosa (Hg.), S. 157-172.
- Klosa, Annette (Hg.) (2011): *elexiko*. Erfahrungsberichte aus der lexikographischen Praxis eines Internetwörterbuchs. (= Studien zur Deutschen Sprache 55). Tübingen: Narr.
- Klosa, Annette (2010): On the combination of automated information and lexicographically interpreted information in two German online dictionaries. In: Granger, Sylviane/Paquot, Magali (Hg.): eLexicography in the 21st century: New challenges, new applications, Louvain-la-Neuve, 22-24 October 2009, Centre for English Corpus Linguistics, Université catholique de Louvain. Louvain-la-Neuve: CECL, S. 157-163.
- Klosa, Annette/Koplenig, Alexander/Töpel, Antje (2011): Benutzerwünsche und Meinungen zu einer optimierten Wörterbuchpräsentation – Ergebnisse einer Onlinebefragung zu *elexiko*. (= OPAL – Online publizierte Arbeiten zur Linguistik 3/2011). Mannheim: Institut für Deutsche Sprache. <http://pub.ids-mannheim.de/laufend/opal/pdf/opal2011-3.pdf>.
- Knowles, Francis E. (1990): The Computer in Lexicography. In: Hausmann et al. (Hg.), S. 1645-1672.
- Landau, Sidney L. (1984): Dictionaries. The Art and Craft of Lexicography. New York: Cambridge University Press.
- Lemberg, Ingrid (2001): Aspekte der Online-Lexikographie für wissenschaftliche Wörterbücher. In: Lemberg/Schröder/Storrer (Hg.), S. 71-91.
- Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (Hg.) (2001): Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. (= Lexikographica. Series Maior 107). Tübingen: Niemeyer.
- Meyer, Peter (2011): *vernetziko*: A Cross-Reference Management Tool for the Lexicographer's Workbench. In: Kosem, Iztok/Kosem, Karmen (Hg.): Electronic lexicography in the 21st Century: New Applications for New Users. Proceedings of eLex2011, Bled, Slowenien, 10-12 November 2011. Ljubljana: Trojina, Institute for Applied Slovene Studies, S. 191-198. www.trojina.si/elex2011/Vsebine/proceedings/eLex2011-25.pdf.

- Meyer, Peter/Müller-Spitzer, Carolin (2010): Consistency of sense relations in a lexicographic context. In: Barbu Mititelu, Verginica/Pekar, Viktor/Barbu, Eduard (Hg.): Proceedings of the Workshop 'Semantic Relations. Theory and Applications', 18 May 2010, at the International Conference on Language Resources and Evaluation (LREC) 2010, Malta. www.lrec-conf.org/proceedings/lrec2010/workshops/W9.pdf.
- Müller-Spitzer, Carolin/Schneider, Roman (2009): Ein XML-basiertes Datenbanksystem für digitale Wörterbücher – Ein Werkstattbericht aus dem Institut für Deutsche Sprache. In: *it-Information Technology* 51.4, S. 197-206.
- Schnörch, Ulrich (2005): Die *elexiko*-Stichwortliste. In: Haß (Hg.), S. 71-90.
- Schröder, Martin (1997): Brauchen wir ein neues Wörterbuchkartell? Zu den Perspektiven einer computerunterstützten Dialektlexikographie und eines Projektes „Deutsches Dialektwörterbuch“. In: *Zeitschrift für Dialektologie und Linguistik* 64, 1, S. 57-65.
- Storjohann, Petra (2011): Paradigmatische Konstruktionen in Theorie, lexikografischer Praxis und im Korpus. In: Klosa (Hg.), S. 99-129.
- Storrer, Angelika (1998): Hypermedia-Wörterbücher: Perspektiven für eine neue Generation elektronischer Wörterbücher. In: Wiegand, Herbert Ernst (Hg.): Wörterbücher in der Diskussion III. (= *Lexikographica. Series Maior* 84). Tübingen: Niemeyer, S. 106-131.
- Storrer, Angelika (2001): Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie. In: Lemberg/Schröder/Storrer (Hg.), S. 54-69.
- Storrer, Angelika/Freese, Katrin (1996): Wörterbücher im Internet. In: *Deutsche Sprache* 24, S. 97-136.
- Wiegand, Herbert Ernst (1998): Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilbd. Berlin/New York: de Gruyter.

4.2 Internetquellen (alle zuletzt eingesehen am 5.2.2015)

- COSMAS II: <https://cosmas2.ids-mannheim.de/cosmas2-web/>
- elexiko* – ein Onlinewörterbuch zur deutschen Gegenwartssprache: www.elexiko.de
- Datenbank gesprochenes Deutsch im Archiv für gesprochenes Deutsch: <http://agd.ids-mannheim.de/datenbanken.shtml>
- Das Deutsche Referenzkorpus DEREKO: www.ids-mannheim.de/kl/projekte/korpora.html
- grammis – das grammatische Informationssystem des IDS: <http://hypermedia.ids-mannheim.de>
- Kookkurrenzdatenbank CCDB – V3.3: <http://corpora.ids-mannheim.de/ccdb/>
- Lexikon zum öffentlichen Sprachgebrauch: www.owid.de/wb/elexiko/projekt/modSprachgeb.html
- Oracle: www.oracle.com/de/products/database/overview/index.html
- Projekt „Kookkurrenzanalyse und deren Erschließung“: www.ids-mannheim.de/kl/projekte/methoden/ka.html
- Projekt „Methoden der Korpusanalyse und -erschließung“, Teilprojekt „Kookkurrenzanalyse und deren Erschließung“: www.ids-mannheim.de/kl/projekte/methoden.html
- Projekt „Texttechnologie und Datenbanken“: www.ids-mannheim.de/gra/projekte/grammis2.html
- Tamino: www.softwareag.com/corporate/products/az/webmethods/default.asp
- Wortartikel *Mobilität* in *elexiko*: www.owid.de/artikel/62377

Sandra Denger, Franziska Horn

Die ersten Schritte eines Langzeitprojekts. Das Digitale Familiennamenwörterbuch Deutschlands

1. Einleitung

Der vorliegende Beitrag beschreibt die ersten Schritte des Digitalen Familiennamenwörterbuch Deutschlands (DFD) im Hinblick auf den lexikographischen Prozess und dessen Phasen.¹

In den ersten Abschnitten des Beitrags wird das DFD mit seiner Organisationsstruktur und den gesetzten Zielen vorgestellt, um dann die zugrunde liegende Motivation und die Notwendigkeit eines solchen Vorhabens zu erläutern. Anschließend wird das Projekt in den aktuellen Forschungszusammenhang eingebettet. Der Schwerpunkt des Artikels beschäftigt sich mit lexikographischen Phasen und diskreten Schritten, die in der Entwicklung und Erstellung eines Wörterbuchs umzusetzen sind. Dies geschieht unter dem Gesichtspunkt der Anforderungen eines Langzeitprojekts, das als Veröffentlichungsmedium das Internet vorsieht. Im Fazit werden aus den bisherigen Erfahrungen der Projektlaufzeit Ergebnisse festgehalten, Schlüsse in Bezug auf medienspezifische Abläufe gezogen und zur Diskussion gestellt. Durch die Beschreibung des Entstehungsprozesses eines digital konzipierten Spezialwörterbuchs soll der Artikel zudem zur Klärung der Kernfrage des Arbeitstreffens – ob und inwiefern sich der lexikographische Prozess durch die Onlinepublikation verändert – beitragen.

2. Projekt

In diesem Kapitel wird das DFD vorgestellt und die Forschungssituation im Bereich der Familiennamenlexikographie beschrieben.

2.1 Organisation und Ziele

Das im Januar 2012 begonnene Akademievorhaben Digitales Familiennamenwörterbuch Deutschlands (DFD) erfolgt in enger Kooperation zwischen der Akademie der Wissenschaften und der Literatur in Mainz, der Johannes Gutenberg-Universität Mainz und der Technischen Universität Darmstadt. Das Langzeitprojekt – mit einer geplanten Laufzeit von 24 Jahren – hat als primäres Ziel, den gesamten Grundbestand der gegenwärtig in Deutschland vorkommenden Familiennamen lexikographisch zu erfassen, zu etymologisieren und zu kartieren. Neben den deutschen werden auch fremdsprachige Familiennamen aufgenommen. Jeder Familienname erhält einen eigenen Artikel im Wörterbuch. Die Etymologie und Erläuterung der Familiennamen steht bei der Erstellung der Namenartikel im Vordergrund. Diese wird durch die erst seit kurzem mögliche flächendeckende Darstellung der geographischen Verbreitung der Familiennamen unterstützt. Dieser neue methodische Zugang bietet den Bearbeiter/-innen

¹ Die Ausführungen stellen den Stand der Arbeiten am DFD vom März 2013 dar. Die inzwischen fortgeschrittene Entwicklung des Redaktionssystems wird bei Denger/Horn (2014) näher thematisiert.

neue Möglichkeiten, die sprachhistorische Herkunft zu klären, zu erläutern bzw. vorhandene Deutungen zu untermauern, zu modifizieren oder zu widerlegen.

In Anbetracht des großen Interesses an Personennamen – ob Ruf- oder Familiennamen – in der Öffentlichkeit ist das DFD auf zwei große Zielgruppen ausgerichtet, die sich für die Bedeutung, Herkunft und Verbreitung von Namen interessieren: Laien und Wissenschaftler/-innen. Letztgenannte entstammen einer Vielzahl an Disziplinen, da die Namenforschung einen interdisziplinären Zugang für verschiedenste Forschungsfragen bereitstellt (vgl. Debus 2012, S. 15). So dienen Familiennamen nicht nur der Sprachwissenschaft oder Dialektologie als Forschungsgegenstand, sondern sie transportieren auch kulturelle Informationen, die Untersuchungsgegenstände für die Kulturgeschichte und Anthropologie darstellen, oder sie dokumentieren Siedlungs- und Wanderbewegungen, welche für die Siedlungs- und Migrationsforschung (vgl. Kunze 1998, S. 187) von Interesse sind. Von einzelnen Orten lässt sich durch die geographische Verteilung der Familiennamen, die nach der Herkunft benannt sind, wie etwa *Nürnberger* oder *Rheinheimer*, beispielsweise der Auszugsbereich lokalisieren (vgl. Kunze 1998, S. 187). Um den unterschiedlichen Anforderungen der Zielgruppen gerecht zu werden, werden neben Angaben zur Etymologie und Verbreitung Sekundärinformationen zur Verfügung gestellt, die zum einen mehr dem Laien – indem etwa ein Glossar und Informationskästen zu speziellen Themengebieten angeboten werden –, zum anderen Wissenschaftler/-innen – die beispielsweise Verweise auf relevante Fachliteratur der Namendeutung erhalten – dienen.

Für die Umsetzung dieser Ziele ist es nötig, sich vorab auf eine gewisse Vorkommenshäufigkeit des Familiennamens innerhalb des Gesamtnameninventars festzulegen. Berücksichtigt werden in der Auswahl primär Familiennamen mit einer Frequenz von 10 Tokens, da dies rechnerisch innerhalb der Projektlaufzeit eine realistisch bearbeitbare Menge an Namen ergibt. Diese Schwelle schließt jedoch zahlreiche niedrigfrequente Namen nicht aus, wenn sie Varianten eines aufgenommenen Familiennamens darstellen. Von den insgesamt ca. 1,1 Millionen verschiedenen Familiennamen (Types) des Korpus – nach Abzug von Bindestrich-Doppelnamen ist mit einem realistischen Inventar von ca. 850.000 Familiennamen zu rechnen – ergibt sich für das DFD ein zu bearbeitendes Nameninventar von knapp über 200.000 Familiennamen.² Jedoch umfasst diese Auswahl selbst nach dem Abzug der niedrigfrequenten Familiennamen noch etwa 95% der Telefonanschlussbesitzer.³

2.2 Motivation

Ein Spezialwörterbuch wie das DFD ist ein dringendes Forschungsdesiderat. Dafür spricht die linguistische Sonderstellung, die Eigennamen einnehmen. Außerdem ist in den Blick zu nehmen, dass allgemeinsprachliche Wörterbücher Eigennamen nur unzulänglich verzeichnen und existierende Nachschlagewerke zu Familiennamen defizitär sind.

Eigennamen nehmen im Sprachsystem vor allem aufgrund ihrer spezifischen Referenzfunktion und ihres grammatischen Verhaltens eine Sonderstellung ein. Im Gegensatz zu den Appellativen, mit denen sie zu den Substantiven gezählt werden, sind Eigennamen monoreferent, d.h. sie referieren idealerweise auf nur ein Denotat (vgl. Nübling et al. 2012, S. 32). Eigenna-

² Aktuelle Lexika umfassen zwischen ca. 20.000 (Duden Familiennamen 2005) und ca. 70.000 (Gottschald 2006) Familiennamen. Das Verzeichnis von Gottschald entspricht damit etwa 8,2% des Gesamtinventars der existenten Familiennamen in Deutschland.

³ Eine detaillierte Projektbeschreibung findet sich bei Fahlbusch/Heuser (2014).

men haben somit die Funktion, ein Individuum zu benennen bzw. zu identifizieren, während Appellative Klassen von Objekten charakterisieren (vgl. Fleischer 1964, S. 377). Des Weiteren unterscheiden sich Eigennamen dadurch von den Appellativen, dass ihnen in der Regel kein lexikalisch-semantischer Inhalt zukommt (vgl. Kunze 1998, S. 11). Da sie sich oft aus Appellativen entwickelt haben, können Eigennamen noch Strukturen lexikalischer Bedeutung enthalten (vgl. Harnisch/Nübling 2004, S. 1901). Diese Konservierung von appellativischem Wortschatz und morphologischen Strukturen betrifft vor allem das relativ junge Nameninventar der Familiennamen (vgl. Marynissen/Nübling 2010, S. 311).

In Anbetracht dieser angesprochenen Sonderstellung stellen Eigennamen auch in der Lexikographie einen besonderen Fall dar und werden deshalb meist in eigenen Nachschlagewerken behandelt. Wenn Eigennamen in allgemeinsprachliche Wörterbücher aufgenommen sind – dies betrifft vor allem Personennamen und geographische Eigennamen (vgl. Klosa et al. 2010, S. 656) –, muss sich folglich sowohl die etymologische Beschreibung als auch die grammatische Erklärung von den anderen Lemmata unterscheiden (vgl. Strauß 1989, S. 788). Die Wörterbuchnutzer/-innen erhalten etwa grammatische Hinweise, wie zur Schreibung und zur Silbentrennung, dabei ist auch hier zu bedenken, dass Eigennamen „eigene grammatische Regeln“ (Nübling et al. 2012, S. 64) befolgen, die in allgemeinsprachlichen Wörterbüchern oft unbehandelt bleiben, so etwa die besondere Pluralflexion.⁴ Die Angabe einer Bedeutung ist anders umzusetzen als bei Appellativen, da Eigennamen synchron keine Bedeutung im appellativischen Sinne haben (vgl. Nübling et al. 2012, S. 32). Die Bedeutungserläuterung des Eigennamens muss über die etymologische Herkunft, das Motiv der Benennung und die Semantik des zugrunde liegenden Etymons erfolgen. Deutlich wird die Problematik durch eine Paraphrase eines Wörterbucheintrags wie zum Beispiel ‘X ist ein deutscher Familienname’. Eine solche Paraphrase bietet die metasprachliche Information, dass es sich bei dem Lemma um einen Familiennamen mit deutschen Wurzeln handelt. Inwieweit den Nachschlagenden damit ein Mehrwert an Information zu ihrem bisherigen Wissen vermittelt wird, ist hinterfragbar, da die Paraphrase zur Bedeutung des Namens nichts beiträgt. Hinzu kommt, dass in der lexikographischen Praxis die Angabe metasprachlichen Wissens in Bezug auf Eigennamen häufig inkonsistent ist. Im Duden (2009) zeigt sich anhand von Stichproben eine unbeständige Handhabung, ob ein Eintrag sprachliches Wissen – „Edelbert (m. Vorname)“ – oder objektbezogenes Wissen – „Mozart (österr. Komponist)“ – zu einem Eigennamen abbildet.⁵

Nicht nur die genannten Probleme von Einträgen zu Eigennamen bzw. Familiennamen in allgemeinsprachlichen Wörterbüchern, sondern auch die Defizite spezieller Nachschlagewerke wie Namen(wörter)bücher, Namenlexika oder Namenatlanten, zeigen, dass die Notwendigkeit eines neu zu erarbeitenden Nachschlagewerks in mehrfacher Hinsicht besteht. Im Vorfeld des Projektantrags wurden bestehende Familiennamenwörterbücher auf verschiedene Kriterien – wie u.a. darin enthaltenes Nameninventar und die Nutzbarkeit – hin geprüft.

Durch diese Studie ist deutlich geworden, dass viele Nachschlagewerke regional begrenzt sind, d.h., dass kaum überregionale Werke zu Familiennamen existieren. Für eine umfassende Namendeutung müssen somit immer mehrere Werke herangezogen werden.

⁴ Die unterschiedliche Strategie der Pluralbildung zwischen Appellativ und Eigenname wird deutlich bei dem Vergleich von Homonymen im Singular. Das Appellativum *Bach* bildet die Pluralform *Bäche*. Der homonyme Familienname *Bach* bildet den Plural *die Bachs* (vgl. Nübling et al. 2012, S. 72).

⁵ Siehe zum Gebrauch metasprachlichen und nichtmetasprachlichen Wissens in Wörterbucheinträgen von Eigennamen auch die Studie von Harweg (1999).

Festzuhalten ist aus dieser Voruntersuchung, dass bislang nur ein Bruchteil der bestehenden Familiennamen lexikographisch fixiert ist und insbesondere Familiennamen mit fremdsprachigen Wurzeln kaum erfasst sind. Seltene Familiennamen, die jedoch die überwiegende Mehrheit des Nameninventars ausmachen, sind nur vereinzelt verzeichnet. Eine Stichprobe zeigt, dass mindestens jeder fünfte der heute vorkommenden Familiennamen in keinem der einschlägigen Nachschlagewerke erfasst ist. Stattdessen sind dort viele heute nicht mehr existierende Familiennamen verzeichnet.

Als weiteres Resultat der Sichtung der Nachschlagewerke ist festzustellen, dass die etymologischen Angaben lücken- und/oder fehlerhaft sind. Gottschald (2006, S. 433) verzeichnet beispielsweise *Schillo* zu sorbisch *šylo* 'Pfriem'. Jedoch ist durch die Verbreitung im Saarland und in der Südpfalz ein Zusammenhang mit dem französischen Patronym *Gillot* (zum Rufnamen *Āgidius* plus Diminutivendung *-ot*) naheliegend. Viele Familiennamen können auch auf verschiedene Wurzeln zurückgeführt werden. Der Familienname *Wiedemann* beispielsweise kann aus *wīde* 'Weide' oder *widem* 'kirchliches Grundstück' (vgl. Kunze 1998, S. 155) hervorgegangen sein. Dieses Spektrum an sogenannten Bedeutungskonkurrenzen wird nur unzureichend angegeben. Durch die neue Methode der Familiennamengeographie lassen sich Bedeutungskonkurrenzen auflösen und dem Familiennamen durch die Verbreitungsanalyse die zutreffende(n) Etymologie(n) zuordnen. Aus heutiger Perspektive unzulängliche oder gar fehlerhafte Angaben in einigen Werken können durch diese neue Methode revidiert, aktualisiert und ergänzt werden.

Da an Familiennamen ein großes öffentliches Interesse besteht, ist auch auf die Nutzbarkeit und die Wissensvermittlung der Nachschlagewerke in Bezug auf Laien geachtet worden. Hier stellt sich heraus, dass viele Nachschlagewerke ohne linguistisches oder sprachhistorisches Vorwissen kaum verständlich sind. Auch sind sie oft für Laien schlecht verfügbar, beispielsweise wegen des Preises oder des Zugangs zu Bibliotheken. Da manche Namen in Sammelartikeln zusammengestellt sind oder die alphabetische Ordnung auf Formen älterer Sprachstufen basiert, erweist sich das Nachschlagen eines Familiennamens oft als Herausforderung. Der Familienname *Robert* ist etwa bei Gottschald (2006) unter dem Eintrag *HROD* verzeichnet, der insgesamt 758 einzelne Namen umfasst, darunter so verschiedene wie *Gruhl*, *Krupp*, *Rhode* oder *Riedel*.

Als Fazit zur gegenwärtigen Situation der lexikographischen Erschließung von Familiennamen lässt sich ziehen, dass für eine umfassende Namendeutung mindestens sechs verschiedene Nachschlagewerke⁶ herangezogen werden müssen, die in vielen Fällen noch um regionale oder spezialisierte – wie für fremdsprachige Namen oder für besondere Namensgruppen – ergänzt werden müssen. Die Erarbeitung eines umfassenden, zentralen, modernen und für jedermann jederzeit zugänglichen Familiennamenwörterbuchs schließt demnach eine empfindliche Forschungslücke. Zu leistende Aufgaben der Bearbeiter/-innen des DFD sind, viele Familiennamen erstmals zu etymologisieren und bestehende Deutungen zusammenzuführen, zu präzisieren und neu zu deuten.

2.3 Aktuelle Forschungslage und Projekte der Namenforschung

Das beschriebene Defizit in der Namenforschung wird aktuell durch mehrere Projekte im europäischen Raum angegangen. Zurzeit entstehen in den Nachbarländern Deutschlands ähnli-

⁶ Bahlw (2005), Brechenmacher (1960-63), Gottschald (2006), Kohlheim/Kohlheim (2005), Naumann (2007), Zoder (1968).

che Projekte zu Namenwörterbüchern und -atlanten, mit denen sich bereits Kooperationen und Synergieeffekte ergeben haben. Das Familiennamenbuch der Schweiz stellt in der Online-Datenbank⁷ die Verbreitung und die Herkunft der Familiennamen in der Schweiz öffentlich zur Verfügung. Auch die niederländischen Familiennamen sind in einer frei zugänglichen Online-Datenbank⁸ mit etymologischen Angaben aufgearbeitet. Diese Daten wurden dem DFD-Projekt freundlicherweise zur Recherche tabellarisch mit XML-Codierung überlassen. Die rezenten Familiennamen Österreichs werden durch das Projekt FamOs⁹ des Instituts für Dialekt- und Namenlexika der Österreichischen Akademie der Wissenschaften etymologisch aufgearbeitet. Am *Laboratoire für luxemburgische Sprach- und Literaturwissenschaft* der Universität Luxemburg werden parallel ein Luxemburgischer Familiennamenatlas (LFA)¹⁰ und ein Namenbuch¹¹ zum Bestand der luxemburgischen Familiennamen erarbeitet, welche auch grenzübergreifende Daten – beispielsweise aus Belgien und Frankreich – einbeziehen. Aus der Sicht des DFD ist vor allem das DFG-Projekt Deutscher Familiennamenatlas (DFA)¹² hervorzuheben, eine Kooperation der Johannes Gutenberg-Universität Mainz und der Albert-Ludwigs-Universität Freiburg. Datengrundlage für den Familiennamenatlas und das Familiennamenwörterbuch bildet das Telefonverzeichnis der Telekom (Stand: 30.06.2005), das von Firmenanschlüssen bereinigt wurde. Aus diesen Daten ist im Rahmen des DFA-Projekts eine Datenbank entstanden, die als Informationen den Familiennamen mit Frequenzangabe und Postleitzahl bereitstellt. Das Kartierungsprogramm, das für das DFA-Projekt von Richard Kunze (tivano Software GmbH)¹³ entwickelt wurde, verwertet diese Informationen zur Erstellung der Verbreitungskarten.¹⁴ Diese Datenbasis macht es möglich, sowohl Einzelnamen als auch Namenbestandteile – beispielsweise alle Namen, die auf *-bauer* enden, abzufragen und deutschlandweit zu kartieren – eine für linguistische Zwecke unentbehrliche Abfragetechnik, die andere Namenprojekte vermissen lassen. Während der Familiennamenatlas vor allem einen namengeographischen Ansatz verfolgt und frequente, großräumige linguistische sowie kulturhistorische Phänomene kartiert (vgl. Kunze/Nübling 2007, S. 131), leistet das Familiennamenwörterbuch die ausführliche Etymologisierung und Erklärung der einzelnen Familiennamen.

3. Lexikographischer Prozess

In diesem Abschnitt werden die ersten Aktivitäten des DFD näher beleuchtet sowie anstehende Aufgaben identifiziert und eingeordnet. Thematisiert werden bereits erledigte bzw. noch zu bearbeitende Aufgaben, gefundene Lösungen und derzeit in der Bearbeitung bzw. in der Erprobung befindliche Schritte. Hervorgehoben werden sollen in den Ausführungen Merkmale des lexikographischen Prozesses, die sich insbesondere aus der Einordnung als Spezialwörterbuch sowie der Länge der Projektlaufzeit ergeben.

Dabei unterteilen wir die Überlegungen in vier zentrale Themenfelder: Konzeption, Daten, Technik und Präsentation. Diese dienen als Schlüsselbegriffe, denen verschiedene Prozesse in

⁷ Familiennamenbuch der Schweiz www.hls-dhs-dss.ch/famn/.

⁸ Niederländische Familiennamendatenbank: www.meertens.knaw.nl/nfb/.

⁹ Familiennamen Österreichs (FamOs): <http://hw.oeaw.ac.at/famos>.

¹⁰ Luxemburgischer Familiennamenatlas (LFA): <http://infolux.uni.lu/familiennamen/>.

¹¹ Gilles/Kollmann/Müller (2016).

¹² Deutscher Familiennamenatlas (DFA): Institut für Geschichtliche Landeskunde an der Universität Mainz e.V. www.iql.uni-mainz.de/forschung/namenforschung.html, an der Universität Freiburg <http://portal.uni-freiburg.de/sdd/forschung/DFA>.

¹³ www.tivano.de

¹⁴ Die Datenbank ist für wissenschaftliche Zwecke zugänglich.

der Wörterbucharbeit zum DFD zugeordnet werden. Die Beschreibungen orientieren sich an den Ausführungen von Klosa in ihrem Aufsatz *The lexicographical process (with special focus on online dictionaries)* (2013), da sie grundlegende Schritte und Prozesse eines dynamischen Online-Wörterbuchs sichtbar werden lässt. Sie identifiziert sechs Phasen, die sich innerhalb des lexikographischen Prozesses auch vermischen und überschneiden können: 1. Vorbereitung, 2. Datensammlung, 3. Technikeinsatz, 4. Datenverarbeitung, 5. Datenanalyse sowie 6. Vorbereitung des Online-Releases.¹⁵ Wie stark die wechselseitigen Bezüge in dieser frühen Projektphase des DFD als einem langfristig angelegten dynamischen Spezialwörterbuch ohne verstetigte Herstellung- und Produktionsstrukturen sind, verdeutlichen die folgenden Ausführungen.

3.1 Konzeption

Zunächst werden Aspekte aufgeführt, die die Organisation des Wörterbuchvorhabens betreffen. Anschließend wird die Konzeption des DFD in Bezug auf Wörterbuchtyp, Makro- und Mikrostruktur thematisiert.

3.1.1 Organisation des Wörterbuchvorhabens

In Vorbereitung des Wörterbuchprojekts galt es, Entwürfe zu entwickeln, Fragen der Finanzierung, des zeitlichen Rahmens sowie des Arbeitsablaufs zu klären. Die Personalplanung schloss mit ein, neben dem lexikographischen Team auch Mitarbeiter/-innen mit Informatik- bzw. EDV-philologischen Kompetenzen zu integrieren. Weiterhin war die Einbindung von externen Fachkräften für die Bearbeitung einzelner Lemmastrecken zu bedenken, die fremdsprachige Namen enthalten. Zu berücksichtigen war auch, dass aufgrund der Einbindung auswärtiger Bearbeiter/-innen und der Kooperation der Akademie in Mainz mit der TU Darmstadt ein dezentrales Arbeiten am Wörterbuch erforderlich sein würde.

Zur Pilotierung des Projekts gehörten u.a. die Festlegung der Stichwörter sowie die Erarbeitung der Informations- bzw. Artikelpositionen. Aus den Überlegungen wurden schrittweise Leitlinien für eine mögliche Artikelbearbeitung gewonnen. Erste Ergebnisse in Bezug auf Konzeption, technische Umsetzung und geplante Nutzungsmöglichkeiten des DFD wurden Fachwissenschaftler/-innen bei einem den Projektantrag vorbereitenden Workshop 2010 vorgestellt und mit ihnen diskutiert (vgl. Ernst 2010).

3.1.2 Wörterbuchtyp

Für eine erste Verortung des DFD mit seinen angedachten Merkmalen in der Wörterbuchlandschaft bietet es sich an, die phänomenologische Typologie¹⁶ von Hausmann (1989) zu nutzen, da sie Bestandteile, Informationspositionen und die Struktur eines Wörterbuches als Kriterien für eine Einordnung heranzieht (vgl. Hausmann 1989, S. 971). Abbildung 1 veranschaulicht diese Typologie und zeichnet die Einstufung des DFD nach.

¹⁵ Eigene Übersetzung der bei Klosa (2013, S. 520) wie folgt benannten Phasen: 1. Preparation, 2. Data Acquisition, 3. Computerization, 4. Data Processing, 5. Data Analysis und 6. Preparation for Online Release. Siehe für eine Beschreibung der Phasen auch den Beitrag von Klosa in diesem Band.

¹⁶ Näheres zum Begriff der Typologie insbesondere in Abgrenzung zu einer Klassifikation findet sich bei Hausmann (1989, S. 968-969).

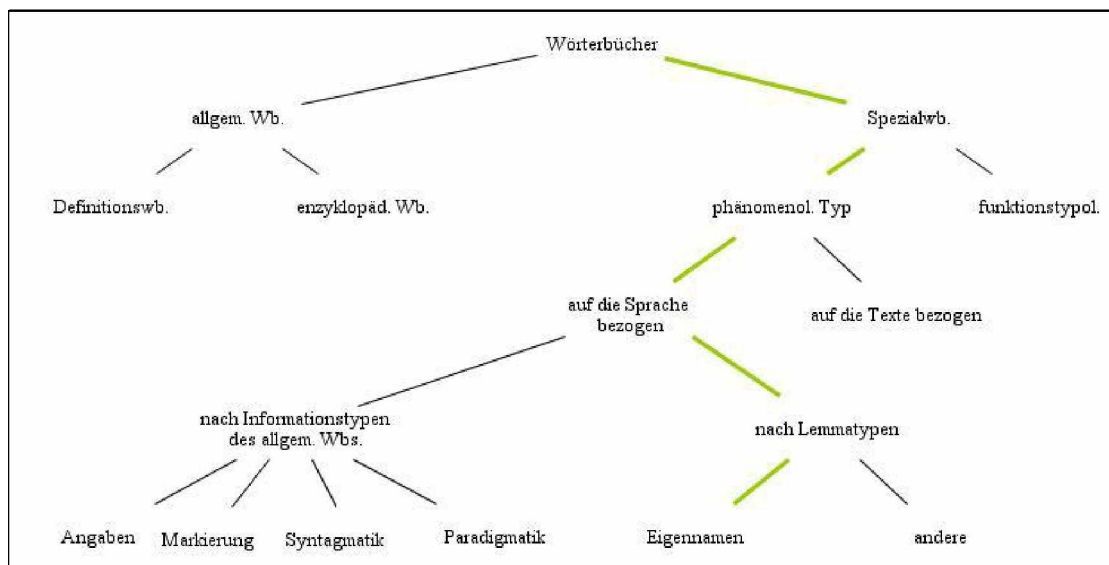


Abb. 1: Einordnung des DFD in die phänomenologische Typologie, eigene Darstellung nach Hausmann (1989, S. 977)

Es handelt sich beim DFD um ein ‘Spezialwörterbuch’, da es sich auf einen Ausschnitt der Sprache konzentriert und die beschriebenen sprachlichen Informationen eingeschränkt sind. Weiterhin erscheint die Einordnung als ‘phänomenologischer Typ’ sinnvoll, da Namen als behandelte Gegenstände im Vordergrund stehen und nicht die konkrete Benutzungssituation, wie etwa bei einem Lernerwörterbuch. Im Gegensatz zu einem Autorenwörterbuch oder einem fiktionalen Namenbuch, das z.B. die literarischen Figuren bei Shakespeare als Untersuchungsgegenstand hat, lässt sich der Gliederungspunkt ‘auf die Sprache bezogen’ zuordnen.

Ferner konzentriert sich das DFD auf die Verzeichnung von bestimmten Lemmatypen, den Familiennamen als Untergruppe der Eigennamen.¹⁷ Eine geographische Einschränkung findet sich, indem in Deutschland vorkommende Familiennamen erklärt werden.

Weitere konzeptionelle Merkmale des DFD lassen sich über das Anwenden alternativer Einteilungsraster herausstellen. Wird die Typologie der Wörterbücher nach Benutzungsmöglichkeiten von Kühn (1989) genutzt,¹⁸ lassen sich verschiedene Verwendungszusammenhänge hervorheben, die, wie in Abbildung 2 dargestellt, die intendierte Funktion des DFD illustrieren:

¹⁷ Beschreibungen zu verschiedenen Formen von Namenwörterbüchern finden sich bei z.B. Engelberg/Lemnitzer (2009, S. 67-68) sowie bei Seibicke (1990a; 1990b), Schmid (1990) und Greule (1990).

¹⁸ Die verschiedenen Benutzungsmöglichkeiten und Schwierigkeiten der Zuordnung von potentiellen Benutzergruppen ohne fundierte empirische Forschungsergebnisse werden detailliert bei Kühn (1989) beschrieben.

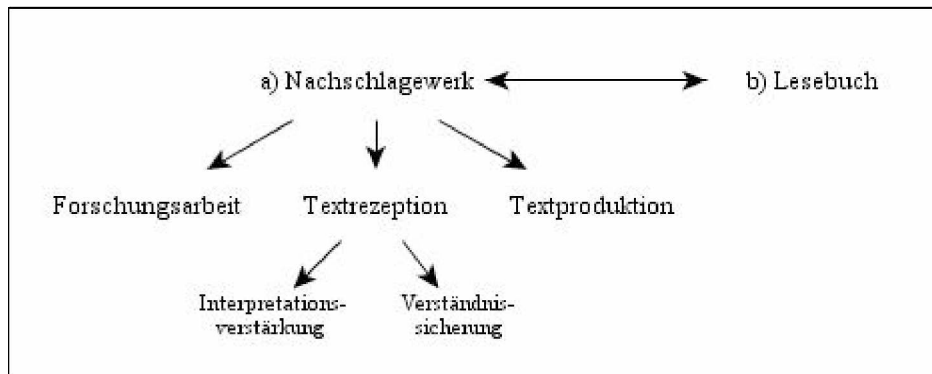


Abb. 2: Benutzungsmöglichkeiten des DFD nach der Typologie von Kühn, eigene Darstellung nach Kühn (1989, S. 121)

a) DFD als Nachschlagewerk

Das DFD kann zur ‘Textrezeption’ und ‘Verständnissicherung’ konsultiert werden. Denkbar wäre auch eine Nutzung zur ‘Interpretationsverstärkung’ z.B. bei in literarischen Texten vorkommenden Familiennamen. Bei der ‘Textproduktion’ erscheint ein Nachschlagen einerseits in Bezug auf komplexe oder fremdsprachige Namen denkbar. Andererseits könnte das Wörterbuch auch Anregung bieten, wenn Namen beim Verfassen von literarischen Texten vergeben werden. Ebenfalls zu erwähnen ist die Verwendung des DFD in Zusammenhang mit ‘Forschungsarbeiten’ verschiedener Disziplinen.

b) DFD als Lesebuch

Zu bedenken ist auch ein möglicher Nutzen des DFD ohne ein unmittelbares Informationsbedürfnis aus einem allgemeineren Interesse am Wörterbuchgegenstand heraus. Einen Anreiz zum Durchstöbern des Wörterbuchs könnten z.B. seltene oder kurios wirkende Namen, Varianten des eigenen Namens und interaktive Verbreitungskarten bieten. Die Verwendung als Lesebuch wird unterstützt durch die Hypertextstruktur des DFD, über die verwandte Namen beispielweise miteinander verlinkt sind.

Wird das DFD nach seiner medialen Umsetzung eingeordnet, lässt es sich wie *ellexiko* als dynamisches Online-Wörterbuch (vgl. Hahn et al. 2008, S. 80; Klosa 2013, S. 519) bezeichnen. Eine Teilpublikation im Printmedium ist jedoch nicht ausgeschlossen und kann zu einem späteren Zeitpunkt erfolgen. Im Hintergrund der verschiedenen Publikationsoptionen steht eine medienneutrale Datenhaltung, die sich über die Verwendung des XML-Formats auszeichnet. Hier zeigt sich, dass die Planung und Konzeption des DFD Aspekte von Langfristigkeit und Nachhaltigkeit berücksichtigt.

Zusammengefasst lässt sich das DFD als ein Spezialwörterbuch einordnen, das als dynamisches Online-Wörterbuch sukzessive die in Deutschland vorkommenden Familiennamen erklärt und für verschiedene Benutzungsmöglichkeiten kostenfrei zugänglich macht.

3.1.3 Makrostruktur

Es wird eine glattalphabetische Anordnung des Namenbestandes angestrebt, sodass wie beim Projekt FamOs (vgl. Linsberger 2011, S. 157) jeder Name, jeweilige graphische Varianten und Komposita einen eigenen Eintrag erhalten. Der Vorteil einer glattalphabetischen Anordnung, wie sie Engelberg/Lemnitzer (2009, S. 257) für elektronische Wörterbücher ohne Platz-

beschränkungen des gedruckten Mediums fordern, wird in der Benutzerfreundlichkeit gesehen: Einzelne Namen sind schneller und leichter auffindbar, was insbesondere auf Varianten von Namen zutrifft, deren Verbindung zur Hauptvariante nicht sofort transparent erscheint. In diesem Vorgehen wird eine Verbesserung gegenüber anderen Namenwörterbüchern, wie z.B. dem Gottschalds (2006) gesehen, der eine Anordnung mit Nischenartikeln verfolgt.

Was die Zugriffsmöglichkeiten betrifft, so wird die strikt alphabetische Ordnung über alternative Zugänge auf den Lemmabestand aufgelöst. Beispielhaft sind einfache und komplexe Suchmöglichkeiten sowie die Anordnung nach zusätzlichen Kategorien (z.B. dem Benennungsmotiv eines Familiennamens) zu erwähnen. Denkbar erscheint in einem Namenwörterbuch auch die Anwahl eines Lemmas über eine Karte (vgl. Winner/Buchner 2011, S. 192).¹⁹

Zugleich soll ein intensives Verweissystem unter Ausnutzung des Hypertextkonzepts (vgl. Storrer 2001) eingerichtet werden. Für den Umgang mit dem online angebotenen Namenmaterial sollen auch Export- und Speichermöglichkeiten bereitgestellt werden. Geplant ist weiterhin die Integration des Wörterbuchs in ein Portal²⁰ zur Namenforschung im deutschsprachigen Raum (namenforschung.net), um Möglichkeiten der Verbindung mit anderen Projekten zu schaffen und weiterführende Informationen zu onomastischen Themen anzubieten.

Entgegen einem alphabetischen Vorgehen wird das Namenmaterial modularisiert bearbeitet (vgl. Klosa 2013, S. 519; Storrer 2001, S. 62-63). Es handelt sich um einen kombinierten Ansatz, wobei Lemmata zum einen primär nach ihrer Frequenz sowie nach inhaltlich-formalen Kriterien ausgesucht werden. Letzteres bedeutet, dass beispielsweise thematische Blöcke (z.B. Namen aus bestimmten Berufsbezeichnungen) gebildet werden. Gleichzeitig werden bei der Bearbeitung von Hauptvarianten ebenso die Varianten abgedeckt (z.B. Simplex *Meyer* und sämtliche der etwa 1.700 *-meyer*-Komposita). Beides entspricht einem ökonomischen Vorgehen, das Synergieeffekte nutzt. Indem die Reihenfolge der Bearbeitung durch die Häufigkeit der Namen mitbestimmt wird, können sich z.B. bereits zu Beginn des Projekts viele Träger/-innen häufiger Namen in Erklärungen wiederfinden. Da die Frequenz jedoch nicht allein ausschlaggebend ist, wird eine gleichmäßige Progression des Projekts gewährleistet.

3.1.4 Mikrostruktur

Im Zusammenhang mit der Mikrostruktur des DFD gilt es herauszustellen, welche Konsequenzen sich aus der Spezialisierung des Wörterbuchs auf Namen ergeben.

Da Namen keine lexikalische Bedeutung besitzen, liegt der Schwerpunkt eines DFD-Artikels auf der Klärung der Etymologie einschließlich der Darstellung möglicher Bedeutungskonkurrenzen und der geographischen Verbreitung eines Familiennamens. Die Namen zu etymologisieren, umfasst auch eine explizite Benennung der jeweiligen Herkunftssprache sowie des Herkunftslandes bzw. der entsprechenden Region, wobei immer auch mehrere Angaben möglich sind. Diese Informationseinheiten haben sich im Verlauf der Konzeption bei der vermehrten Erprobung der Artikelerstellung als wichtig herausgestellt, um die Herkunft der fremdsprachigen Namen hervorzuheben. Die Verbreitung des Familiennamens als Bestandteil eines

¹⁹ Beispielhaft ist eine räumliche Zugriffsstruktur bei dem Projekt Ortsnamen.ch umgesetzt: www.ortsnamen.ch. Vorteile und Möglichkeiten, die die Erschließung lexikographischer Informationen über Karten bietet, werden für Dialektwörterbücher bei Wandl-Vogt (2006) erläutert.

²⁰ Grundsätzliche Überlegungen zur Konzeption von Portalen, die lexikographische Inhalte bündeln, finden sich bei Klosa/Lemmitzer/Neumann (2008) sowie bei Engelberg/Lemmitzer (2009, S. 73-81).

Artikels umfasst sein Vorkommen innerhalb und außerhalb Deutschlands sowie seine historische Verbreitung. Ursprüngliche Überlegungen wurden um die letztgenannte fakultative Artikelposition ergänzt, da die Angabe bei Unterschieden zu der aktuell beobachtbaren Streuung in Deutschland relevant sein kann: Bei Namen, deren Herkunft in Schlesien oder Westpreußen zu verorten ist, zeigen aktuelle Verbreitungskarten in Deutschland meist nur ein verwirrendes Streubild.

Die Verbreitung eines Namens wird über Karten veranschaulicht, wobei diese jedoch nicht nur eine illustrierende Funktion einnehmen, sondern deutungsunterstützend wirken (vgl. Nübling/Kunze 2005, S. 141-142). Sich Methoden der computergestützten Familiennamengeographie zu bedienen, bedeutet, eine verhältnismäßig neue Methode der Namenforschung anzuwenden (vgl. Rampl 2011, S. 144). Eine weitere Informationsposition nimmt die Ausspracheangabe ein, die z.B. bei komplex gebildeten Namen bzw. fremdsprachigen Namen für Wörterbuchnutzer/-innen interessant sein kann. Ferner können die Artikel historische Belege enthalten. Der Fokus liegt indessen nicht auf grammatischen Angaben. Zu einem Lemma können entsprechende Varianten in einer Auswahl aufgeführt werden. Dieses Vorgehen ähnelt Bestrebungen des österreichischen Projekts FamOs, Schreibvarianten des Lemmas in den Artikeln zu vermerken (vgl. Linsberger 2011, S. 158). Informationskästen können als zusätzliches Informationsangebot verschiedene Themen wie z.B. die Namenvergabe in anderen Ländern und Kulturen ausführlicher darstellen.

Geplant ist weiterhin die Aufnahme von Literaturangaben in den etymologischen Erläuterungen. Deutungen zu belegen, erscheint insbesondere relevant, da auch Erklärungen in die Artikel aufgenommen werden, die in anderen Nachschlagewerken verzeichnet, aber nach neusten Erkenntnissen unwahrscheinlich sind. In der Auseinandersetzung mit früheren und teilweise nicht zutreffenden Deutungen gilt es folglich, Nachvollziehbarkeit der Ausführungen zu gewährleisten. In einem Artikel verwendete, erklärungsbedürftige Termini sind in einem Glossar allgemein verständlich aufbereitet.

In Bezug auf Literaturangaben, Informationskästen und eine umfangreiche Aufnahme von onymischen Erklärungsansätzen lässt sich der Vorteil einer Online-Publikation hervorheben, die im Gegensatz zu einer Printpublikation keinen Platzbeschränkungen ausgesetzt ist. Als zusätzliches Informationsangebot ist es geplant, unter Ausnutzung des Hypertextkonzeptes auch weiterführende Links anzubieten.

Insgesamt lässt sich anführen, dass die Unterschiedlichkeit von Namen und individuelle Besonderheiten in verschiedenen Kultur- bzw. Sprachräumen eine Herausforderung bei der Erarbeitung einer einheitlichen Mikrostruktur für die Wörterbucheinträge darstellen. Deshalb ist es notwendig, Richtlinien auf der Basis möglichst verschiedener Musterartikel zu erarbeiten.

Bei der Darstellung der Einträge ist angedacht, auf Mittel der Textverdichtung wie z.B. Abkürzungen zu verzichten. Dieses Vorgehen ist zum einen mit dem Bemühen um Verständlichkeit bzw. Benutzerfreundlichkeit zu begründen (vgl. Herbst/Klotz 2003, S. 82) sowie zum anderen als Konsequenz der quasi unbeschränkten Platzmöglichkeiten online zu sehen (vgl. Storrer 2001, S. 56-57).

Die Planung des DFD in organisatorischer und konzeptioneller Hinsicht schließt Überlegungen der technischen Realisierung mit ein. Dazu gehört die Berücksichtigung der Möglichkei-

ten des Hypertextkonzepts: angestrebte Recherchemöglichkeiten, Realisierung der Verlinkung zu relevanten Websites, Verknüpfung zwischen Varianten und mit allgemeinsprachlichen Wörterbüchern, wie mit *ellexiko* geplant, sowie die Entwicklung eines verschiedenen onomastischen Projekte und Inhalte umfassenden Namen-Portals.

Die erwähnten Arbeitsschritte und Überlegungen sind Bestandteile einer vorbereitenden Phase eines Wörterbuchvorhabens, die den Beginn des lexikographischen Prozesses markiert. In Bezug auf die frühe Projektphase des DFD lässt sich jedoch anmerken, dass nachgelagerte Abläufe wie die Datenverarbeitung und -analyse nicht nur aufbauend folgen, sondern gleichzeitig verlaufen, da die Erarbeitung der Wörterbuchinhalte als Erprobung der konzeptuellen Inhalte verstanden werden kann. Hinzuzufügen ist weiterhin, dass in einer späteren Projektphase bei verstetigten Redaktions- und Publikationsstrukturen Fragen der Konzeption erneut aufkommen können, wenn z.B. notwendige Korrekturen und Ergänzungen im Artikelaufbau erforderlich sind (vgl. Hahn et al. 2008, S. 82).

3.2 Daten

In diesem Abschnitt werden verwendete Daten und Ressourcen vorgestellt, die zum einen Grundlage des Wörterbuches sind sowie zur Erarbeitung herangezogen werden. Die Datenbasis, die als Primärquelle des Wörterbuches bezeichnet werden kann (vgl. Klosa 2013, S. 520), bilden die von der Telekom bereitgestellten Angaben zu den Telefonanschlüssen in Deutschland von 2005. Diese auch für die Erstellung des DFA genutzten Daten liegen in einer Datenbank vor, die Auskunft über einzelne Namen bzw. Namensets und ihre Häufigkeiten erteilt und über ein integriertes Kartierungsprogramm deren Verbreitung anzeigt.

Für die lexikographische Arbeit sind weiterhin Sekundärquellen zu sichten und bereitzustellen (vgl. ebd., S. 520-521), worunter z.B. Nachschlagewerke wie andere Namenwörterbücher, allgemeine und historische Wörterbücher, Dialektwörterbücher, Belegwörterbücher und Heiligennamenlexika fallen, wie auch Grammatiken und Sprachgeschichten. Außerdem gilt es, Portale und Kartierungsprogramme zur Ermittlung der Verbreitung von Namen außerhalb Deutschlands und zur historischen Kartierung zu verzeichnen. Allgemein verwendete Informationen werden zentral in Literatur- und Linklisten gesammelt und gegebenenfalls aktualisiert.

In Bezug auf die Gewinnung, Sichtung und Sondierung relevanter Ressourcen und Daten für das DFD lassen sich die umfangreichen Vorarbeiten für den DFA als vorteilhaft erwähnen. Während die Artikelposition der geographischen Verbreitung und somit die Einbindung entsprechender Karten von Beginn des lexikographischen Prozesses an einen essentiellen Bestandteil der Planungen und ersten Umsetzungsschritte darstellt, werden die Ausspracheangaben, die über Audiodateien realisiert werden könnten, zunächst zurückgestellt. Die umfangreiche Aufbereitung und Verfügbarmachung von Primär- und Sekundärquellen, die zur Phase der Datensammlung zu zählen sind (vgl. ebd.), bilden die Voraussetzung für die Erstellung der Namenartikel.

3.3 Technik

Dem DFD steht mit der DFA-Datenbank bereits ein Recherche- und Abfragetool zur Verfügung. Zu leisten ist die Entwicklung eines Redaktionssystems zum Verfassen der Artikel. Es wird eine digitale Arbeitsumgebung angestrebt, die ortsunabhängiges Arbeiten am DFD ermöglicht und sich an modernen Standards zur elektronischen Texterfassung orientiert. Als Datenformat dient daher XML und es wird wie beim DWDS (vgl. Geyken 2011, S. 12) und DONBO (vgl. Winner/Buchner 2011, S. 186-187) eine Kodierung nach Richtlinien der Text Encoding Initiative (TEI) verfolgt. Insbesondere in Bezug auf die lange Projektlaufzeit und im Bestreben um Langzeitverfügbarkeit der Daten erscheint die Orientierung an de-facto-Standards (vgl. Jannidis 2009, S. 258) in den digitalen Geisteswissenschaften angemessen.

Bei der Einrichtung eines Redaktionssystems wird eine Lösung favorisiert, bei der die Dateneingabe allerdings nicht direkt im XML-Code erfolgt. Stattdessen wird ein Editor-Interface angestrebt, das eine benutzerfreundliche Oberfläche mit einer WYSIWYG-(what you see is what you get)-Ansicht zur Visualisierung des Markups bietet. Entwicklungen im Bereich der EDV-Philologie und der Informatik orientieren sich hierbei an den Bedürfnissen und Kenntnissen innerhalb des lexikographischen Teams. Abbildung 3 veranschaulicht Bisheriges und Geplantes:

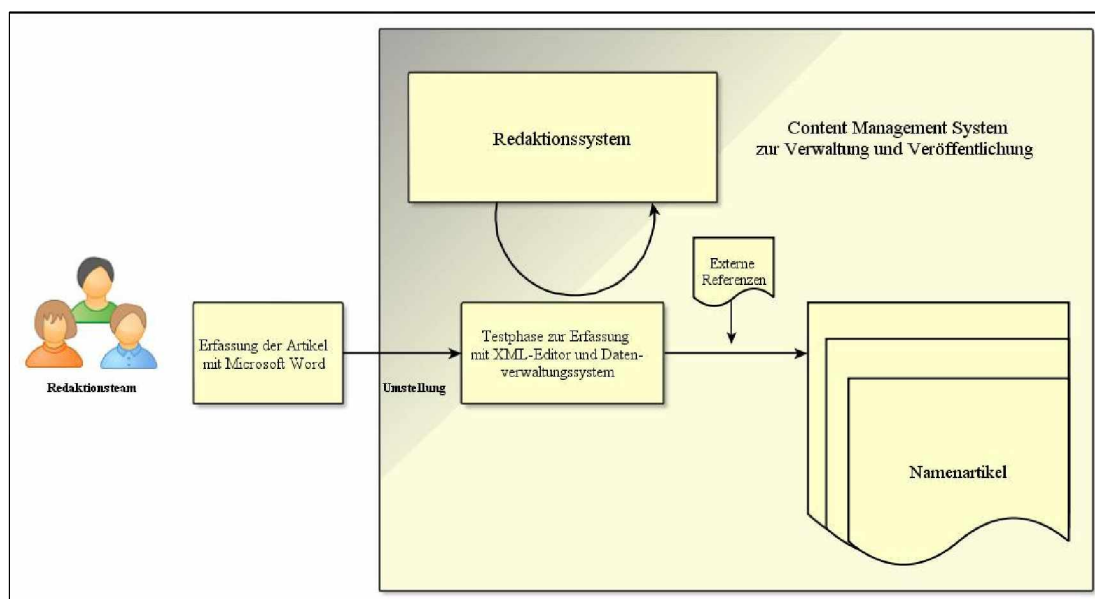


Abb. 3: Technikeinsatz in der frühen Projektphase des DFD, eigene Darstellung²¹

Es gibt eine mehrmonatige Testphase der Artikelerfassung über den auch bei elexiko (Abel/Klosa 2012, S. 416) und dem DWDS (Geyken 2011) genutzten XML-Editor oXygen²² mit einer angepassten Autorenumgebung. Die Umstellung löst die bisherige provisorische Erfassung von Artikeln und Artikelentwürfen über Microsoft Word ab.

Den Bearbeiter/-innen werden Grundgerüstartikel zur Verfügung gestellt, die bereits entsprechend den TEI-Richtlinien aufgebaut sind und Angaben enthalten, die aus der Datenbank automatisch zugewiesen werden können. Dazu zählen die jeweilige Frequenz eines Namens und

²¹ Die Abbildung wurde mithilfe der Software yEd hergestellt: www.yworks.com/de/products_yed_about.html.

²² Es wird die oXygen-Version 14.1. verwendet.

der Rang innerhalb der Familiennamen in Deutschland. Automatisch generiert und im Artikel hinterlegt werden soll für jedes Lemma auch eine ID (identifizier), anhand deren eindeutig und dauerhaft auf die Wörterbuch-Einträge verwiesen werden kann. Die Digitalisierung von Sekundärquellen ist geplant, um Nachschlagewerke zentral und durchsuchbar elektronisch zur Verfügung zu stellen, was einer Arbeitserleichterung für die Artikelerstellung gleichkäme.

Zu den konkreten Arbeitsschritten der lexikographischen Beschreibung gehören z.B. die Kartierung eines Namens sowie die Ermittlung möglicher Varianten des Lemmas. Weiterhin können die Auswertung von Sekundärliteratur und der Vergleich bisheriger Deutungen zur Erhellung der Herkunft eines Namens als wichtige Tätigkeiten herausgegriffen werden. Bei der Artikelerstellung werden externe Referenzen – beispielsweise Kartendaten, weiterführende Internetseiten und verwandte Namen – einbezogen, die verlinkt werden.

Die Ablage und Verwaltung der entstandenen Artikel soll datenbankbasiert erfolgen. Über ein Content-Management-System ist dann die Veröffentlichung und Präsentation geplant. Zu klären sind in diesem Zusammenhang auch Fragen der Versionierung bei Änderungen in einem Artikel und die Langzeitarchivierung.

Über die Testphase wird einerseits beabsichtigt, Einsichten für das Redaktionssystem zu gewinnen. Als weitere Nahziele im lexikographischen Prozess sind andererseits die Verstetigung der Artikelrichtlinien sowie des XML-Schemas über die praktische Erprobung zu nennen.

Dem Bereich Technik lassen sich folglich Aspekte verschiedener Phasen des lexikographischen Prozesses nach Klosa (2013) zurechnen: Die geschilderten Entwicklungen lassen sich der Phase 3, dem Technikeinsatz, aber darüber hinausgehend auch den Phasen 2, 4 und 5, der Datensammlung, -verarbeitung und -analyse, zuordnen. Gleichzeitig findet sich die vorgelagerte Vorbereitungsphase wieder. Die starke Vermischung der einzelnen Abschnitte kann auf die frühe Projektphase zurückgeführt werden, da sich der Einsatz der Technik sowie die konkreten Artikelrichtlinien in der steten Entwicklung und Erprobung befinden. Erkenntnisse aus dem praktischen Einsatz können konzeptuelle Änderungen bewirken, die wiederum Anpassungen der Redaktionsumgebung und Datenhaltung sowie -verarbeitung notwendig machen können.

3.4 Präsentation

Zu diesem Themenbereich gehören für uns Prozesse, die die Veröffentlichung und Online-Stellung des Wörterbuchs betreffen bzw. dorthin führen. Zu nennen ist beispielsweise die Korrektur entstandener Artikelfassungen. Weiterhin sind die Fortführung der Website-Konzeption und des Designs sowie die Entwicklung eines geeigneten Systems zur Datenverwaltung zu erwähnen, was der Vorbereitung des Publikationssystems entspricht.

Im Vorfeld der Online-Veröffentlichung weist Klosa (2013, S. 521) auf die Wichtigkeit des Testens der Artikel in einer simulierten Umgebung hin. So kann die Richtigkeit von Verlinkungen und die korrekte Einbindung der Karten geprüft werden. Weiterhin ist es geplant, die Benutzerfreundlichkeit des Wörterbuchs durch verschiedene Nutzergruppen zu erproben. Zu einer Präsentation online gehören auch Metatexte (vgl. Klosa 2013, S. 521; Engelberg/Lemnitzer 2009, S. 140-141), die Hinweise zur Benutzung des Wörterbuchs geben.

Da es sich beim DFD um ein dynamisches Wörterbuch handelt, wird es ein stetiges Verfassen und Publizieren von Artikeln geben. Dies beinhaltet auch Fragen der Archivierung und Versionierung bei Überarbeitungen bzw. Aktualisierungen eines Artikels (vgl. Klosa 2013,

S. 523). Perspektivisch gesehen sollen auch weitere Projekte im Portal namenforschung.net eingebunden werden. Eine Teilpublikation im Printmedium ist zu einem späteren Zeitpunkt möglich.

Das Feld Datenpräsentation und -veröffentlichung berührt die von Klosa angeführte Phase der Vorbereitung des Online-Releases. Bei möglichen Anpassungen der Internet-Präsentation an Rückmeldungen von Benutzer/-innen ergeben sich Berührungspunkte mit der Konzeption (vgl. Hahn et al. 2008, S. 82). In der frühen Projektphase des DFD bedeutet der Bereich jedoch vor allem konzeptionelle Planung einschließlich Überlegungen zur technischen Umsetzung. Dies deutet eine Überschneidung mit der ersten Phase der Vorbereitung eines Wörterbuchvorhabens und den Phasen der technischen Realisierung an.

Bei verstetigten Redaktionsabläufen in einer späteren Projektphase ist davon auszugehen, dass sich zudem einzelne Phasen in kleinerem Rahmen wiederholen und kreuzen: Im DFD als einem dynamischen Wörterbuch können einzelne Namenstrecken veröffentlicht werden, während andere sich in der Bearbeitung befinden bzw. eine Aktualisierung der Nachschlagewerke erfordern (vgl. Klosa 2013, S. 522). Die Frage der zur Verfügung stehenden Daten kann erneut relevant werden, wenn z.B. neue Namenstrecken wie griechische Namen in Angriff genommen werden.

Zudem können sich konkrete Lemmata innerhalb eines zusammenhängenden Sets in unterschiedlichen Phasen befinden (vgl. Klosa 2013, S. 522). Es ist denkbar, dass Bedeutungskonkurrenzen für eine Namensvariante noch gewichtet werden, während der Eintrag für die dazugehörige Hauptvariante bereits korrekturgelesen wird. Zudem können sich Phasenabläufe zum Teil wiederholen, wenn bereits veröffentlichte Artikel aufgrund von Änderungen aktualisiert und somit in gewisser Hinsicht erneut publiziert werden.

Die Planungen und die Durchführung einer möglichen Printpublikation erfordern aufbauend auf dem medienneutral konzipierten lexikographischen Prozess der Online-Ausgabe ferner erneute, veränderte, konzeptionelle Überlegungen in Anpassung an das veränderte Medium, z.B. über Auswahl des veröffentlichten Materials, den technischen Einsatz zur Aufbereitung und die explizite Veröffentlichung.

4. Fazit

Rückblickend betrachtet ist für die ersten Schritte des DFD festzuhalten, dass sich die Arbeitsschritte weitgehend auf konzeptioneller Ebene befinden, indem sie etwa auf die Ausarbeitung der Artikelpositionen, die technischen Anforderungen und deren Umsetzung bezogen sind. Anknüpfend gilt es, die Konzepte auf ihre Praktikabilität hin zu testen und Modifizierungen vorzunehmen.

In die Planung und Vorbereitung des Wörterbuchprojekts spielen bereits vermehrt weitere Phasen des bei Klosa (2013) beschriebenen lexikographischen Prozesses hinein. Die einzelnen Phasen zur Entwicklung und Erstellung des Wörterbuchs greifen aktuell stark ineinander und verlaufen zum Teil auch zirkulär und nicht strikt nacheinander ab. Es ergibt sich eine Wechselwirkung zwischen konzeptionellen Phasen der Planung und eigentlich nachgelagerten Phasen der konkreten Umsetzung. Deshalb muss eine stetige Evaluation von einzelnen Arbeitsschritten erfolgen. In der Folge können eventuell Modifikationen vorgenommen werden, die

auch auf die Planungs- bzw. Entwicklungsphase zurückwirken können. Dies kann durch die vielfältigen und speziellen Namenstrecken auch den Artikelaufbau betreffen und somit die Textauszeichnung, wobei angestrebt ist, eine offene, allgemein nutz- und handhabbare Artikelvorlage als Eingabemaske zu entwickeln.

Eine Herausforderung für die Planung und Entwicklung des Wörterbuchs stellt zudem die lange Projektlaufzeit dar. Dies setzt bereits in dieser frühen Projektphase vorausschauendes Planen insbesondere in Bezug auf die EDV und Informatik voraus. Dies wird u.a. daran deutlich, dass bereits in dieser frühen Phase des Projekts das Thema Versionierung eine zentrale Rolle in der Planung spielt und mögliche Lösungen zur Archivierung erprobt werden. Aufgrund der langen Laufzeit ist auch davon auszugehen, dass Mitarbeiter/-innen wechseln. In Bezug auf die Texterfassung ist deshalb ein einfacher Zugang zur digitalen Arbeitsumgebung erforderlich. Des Weiteren ist eine ausführliche Dokumentation der Arbeitsschritte zur Artikelerstellung unerlässlich, um das Vorgehen für neue Mitarbeiter/-innen nachvollziehbar zu machen und Entscheidungen auch für spätere Projektphasen transparent zu halten.

Zentrale Fragen und Themen, die den gesamten lexikographischen Prozess des DFD begleiten und überspannen, betreffen, neben den inhaltlichen Komponenten (wie Namendeutung, Arbeitsschritte zur Erstellung des Artikels, Artikelpositionen, Richtlinien zur Bearbeitung bzw. Texterstellung etc.) eben auch vermehrt technische Komponenten. Hierzu gehören allgemeine Fragen und Lösungen zur Langzeitarchivierung sowie Hypertextkonzepte zur Vernetzung innerhalb des Wörterbuchs sowie Vernetzungen zu anderen Projekten und Websites. Doch auch projektintern ist von Seiten der Technik stets darauf zu achten, dass die technischen Umsetzungen innerhalb verschiedener lexikographischer Phasen für die Bearbeiter/-innen nutzbar und nützlich sind. Für das Gelingen des Online-Wörterbuchs ist eine enge interdisziplinäre Zusammenarbeit und Kommunikation unabdingbar, um lexikographische, onomastische, informatische und EDV-philologische Kompetenzen zusammenzuführen.

5. Literatur

- Abel, Andrea/Klosa, Annette (2012): Der lexikographische Arbeitsplatz – Theorie und Praxis. In: Vatvedt Fjeld, Ruth/Torjusen, Julie Matilde (Hg.): Proceedings of the 15th EURALEX International Congress in Oslo 2012, S. 413-421.
- Bahlow, Hans (2005): Deutsches Namenlexikon. Familien- und Vornamen nach Ursprung und Sinn erklärt. 16. Aufl. Frankfurt a.M.
- Brechenmacher, Josef Karlmann (1960-63): Etymologisches Wörterbuch der deutschen Familiennamen. 2 Bde. Limburg.
- Debus, Friedhelm (2012): Namenkunde und Namengeschichte. Eine Einführung. Berlin.
- Denzler, Sandra/Horn, Franziska (2014): Die Arbeitsumgebung des Digitalen Familiennamenwörterbuchs Deutschlands. Ein XML-basiertes Redaktionssystem. In: Mann, Michael (Hg.): Digitale Lexikographie. Ein- und mehrsprachige elektronische Wörterbücher mit Deutsch: aktuelle Entwicklungen und Analysen. (= Germanistische Linguistik 223-224). Hildesheim/Zürich/New York, S. 67-96.
- Duden – Die deutsche Rechtschreibung (2009). 25. Aufl. Mannheim.
- Engelberg, Stefan/Lemnitzer, Lothar (2009): Lexikographie und Wörterbuchbenutzung. 4., überarb. u. erw. Aufl. Tübingen.
- Ernst, Peter (2010): Bericht über den Workshop „Digitales Familiennamenwörterbuch Deutschlands“ (DFD). In: Eichler, Ernst/Hengst, Karlheinz/Kremer, Dietlind: Namenkundliche Informationen 98. Leipzig, S. 245-248.
- Fahlbusch, Fabian/Heuser, Rita (2014): Das „Digitale Familiennamenwörterbuch Deutschlands“. Möglichkeiten und Ziele am Beispiel regionaler Namen. In: Gilles, Peter/Kollmann, Cristian/Müller, Claire (Hg.): Familiennamen zwischen Maas und Rhein. Frankfurt a.M. u.a., S. 209-226.

- Fleischer, Wolfgang (1964): Zum Verhältnis von Name und Appellativum im Deutschen. In: Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig 13, S. 369-378.
- Geyken, Alexander (2011): Die dynamische Verknüpfung von Kollokationen mit Korpusbelegen und deren Repräsentation im DWDS-Wörterbuch. In: Klosa, Annette/Müller-Spitzer, Carolin (Hg.): Datenmodellierung für Internetwörterbücher. 1. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“. (= OPAL – Online publizierte Arbeiten zur Linguistik 2/2011). Mannheim, S. 9-20. <http://pub.ids-mannheim.de/laufend/opal/pdf/opal2011-2.pdf>.
- Gilles, Peter/Kollmann, Christian/Müller, Claire (2016): Luxemburger Familiennamenbuch. Berlin/New York.
- Gottschald, Max (2006): Deutsche Namenkunde. 6. Aufl. Berlin/New York.
- Greule, Albrecht (1990): Ortsnamenwörterbücher. In: Hausmann et al. (Hg.), S. 1276-1284.
- Hahn, Marion/Klosa, Annette/Müller-Spitzer, Carolin/Schnörch, Ulrich/Storjohann, Petra (2008): *elexiko* – das elektronische, lexikografisch-lexikologische korpusbasierte Wortschatzinformationssystem. Zur Neukonzeption, Erweiterung und Revision einzelner Angabebereiche. In: Klosa (Hg.), S. 57-86. http://pub.ids-mannheim.de/laufend/opal/pdf/opal08-1_hahn-klosa-mueller-spitzer.pdf.
- Hamisch, Rüdiger/Nübling, Damaris (2004): Namenkunde. In: Booij, Geert/Lehmann, Christian/Mugdan, Joachim (Hg.): Morphologie / Morphology. Ein internationales Handbuch zur Flexion und Wortbildung / An International Handbook on Inflection and Word-Formation. 2. Halbbd. (= Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 17.2) Berlin/New York, S. 1901-1910.
- Harweg, Roland (1999): Studien zu Eigennamen: Eigennamen als Einträge in Wörterbüchern und Lexika. Aachen, S. 61-123.
- Hausmann, Franz Josef (1989): Wörterbuchtypologie. In: Hausmann et al. (Hg.), S. 968-980.
- Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (Hg.) (1989): Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie. 1. Teilbd. (= Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 5.1). Berlin/New York.
- Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (Hg.) (1990): Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie. 2. Teilbd. (= Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 5.2). Berlin/New York.
- Herbst, Thomas/Klotz, Michael (2003): Lexikografie. Paderborn.
- Jannidis, Fortis (2009): TEI in a crystal ball. In: Literary and Linguistic Computing 24.3, S. 253-265.
- Klosa, Annette (2013): The lexicographical process (with special focus on online dictionaries). In: Gouws, Rufus/Hjalmar/Heid, Ulrich/Schweickard, Wolfgang/Wiegand, Herbert Ernst (Hg.): Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. (= Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 5.4). Berlin/Boston, S. 517-524.
- Klosa, Annette (Hg.) (2008): Lexikografische Portale im Internet. (= OPAL – Online publizierte Arbeiten zur Linguistik. Sonderheft 1/2008). Mannheim. <http://pub.ids-mannheim.de/laufend/opal/pdf/opal2008-1.pdf>.
- Klosa, Annette/Lemnitzer, Lothar/Neumann, Gerald (2008): Wörterbuchportale – Fragen der Benutzerführung. In: Klosa (Hg.), S. 5-36. http://pub.ids-mannheim.de/laufend/opal/pdf/opal08-1_klosa-lemnitzer-neumann.pdf.
- Klosa, Annette/Schnörch, Ulrich/Schoolaert, Sabine (2010): Stichwort, Stichwortliste und Eigennamen in *elexiko*: Einflüsse der Korpusbasiertheit und Hypermedialität auf die lexikografische Konzeption. In: Dykstra, Anne/Schoonheim, Tanneke (Hg.): Proceedings of the XIV Euralex International Congress. Leeuwarden, 6-10 Juli 2010. Leeuwarden, S. 653-663.
- Kohlheim, Rosa/Kohlheim, Volker (2005): Duden Familiennamen. Herkunft und Bedeutung. Mannheim.
- Kunze, Konrad (1998): dtv-Atlas Namenkunde. Vor- und Familiennamen im deutschen Sprachgebiet. München.
- Kunze, Konrad/Nübling, Damaris (2007): Der deutsche Familiennamenatlas (DFA). Konzept, Konturen, Kartenbeispiele. In: Bergmann, Rolf/Nübling, Damaris/Obst, Ulrich/Tiefenbach, Heinrich/Untermann, Jürgen (Hg.): Beiträge zur Namenforschung 42, 2, S. 125-172.
- Kühn, Peter (1989): Typologie der Wörterbücher nach Benutzungsmöglichkeiten. In: Hausmann et al. (Hg.), S. 111-127.

- Linsberger, Axel (2011): Regionalspezifisch als Kriterium der Namensauswahl bei der Erstellung eines Österreichischen Online-Familiennamensbuches. In: Ziegler, Arne/Windberger-Heidenkummer, Erika (Hg.): Methoden der Namensforschung. Methodologie, Methodik und Praxis. Berlin, S. 155-167.
- Naumann, Horst (2007): Das große Buch der Familiennamen. Augsburg/München.
- Nübling, Damaris/Kunze Konrad (2005): Familiennamenforschung morgen: Der Deutsche Familiennamenatlas (DFA). In: Brendler, Andrea/Brendler, Silvio (Hg.): Namensforschung morgen: Ideen, Perspektiven, Visionen. Hamburg, S. 141-151.
- Nübling, Damaris/Fahlbusch, Fabian/Heuser, Rita (2012): Namen. Eine Einführung in die Onomastik. Tübingen.
- Marynissen, Ann/Nübling, Damaris (2010): Familiennamen in Flandern, den Niederlanden und Deutschland. Ein diachroner und synchroner Vergleich. In: Dammel, Antje/Kürschner, Sebastian/Nübling, Damaris (Hg.): Kontrastive germanistische Linguistik. (= Themenband in der Reihe „Germanistische Linguistik 206-209). Hildesheim/Zürich/New York, S. 311-364.
- Rampl, Gerhard (2011): Belegverortung mit GIS: Methodische Aspekte und Anwendungsmöglichkeiten. In: Ziegler, Arne/Windberger-Heidenkummer, Erika (Hg.): Methoden der Namensforschung. Methodologie, Methodik und Praxis. Berlin, S. 143-154.
- Schmid, Wolfgang Paul (1990): Gewässernamenwörterbücher. In: Hausmann et al. (Hg.), S. 1284-1291.
- Seibicke, Wilfried (1990a): Personennamenwörterbücher. In: Hausmann et al. (Hg.), S. 1267-1275.
- Seibicke, Wilfried (1990b): Weitere Typen des Namenwörterbuchs. In: Hausmann et al. (Hg.), S. 1291-1296.
- Storrer, Angelika (2001): Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie. In: Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (Hg.): Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. (= Lexikographica. Series Maior 107). Tübingen, S. 53-69.
- Strauß, Gerhard (1989): Angabe traditioneller Wortarten oder Beschreibung nach funktionalen Wortklassen im allgemeinen einsprachigen Wörterbuch? In: Hausmann et al. (Hg.), S. 788-796.
- Wandl-Vogt, Eveline (2006): Von der Karte zum Wörterbuch – Überlegungen zu einer räumlichen Zugriffsstruktur für Dialektwörterbücher. Dargestellt am Beispiel des Wörterbuchs der bairischen Mundarten in Österreich (WBÖ). In: Corino, Elisa/Marello, Carla/Onesti, Cristina (Hg.): Proceedings of the XII EURALEX International Congress in Torino 2006. Alessandria, S. 721-732.
- Winner, Sabina/Buchner, Martina (2011): Digitales Ortsnamenbuch Online (DONBO): Neue Perspektiven der Namensforschung. In: Ziegler, Arne/Windberger-Heidenkummer, Erika (Hg.): Methoden der Namensforschung. Methodologie, Methodik und Praxis. Berlin, S. 183-198.
- Zoder, Rudolf (1968): Familiennamen in Ostfalen. 2 Bde. Hildesheim.

Christian M. Meyer, Iryna Gurevych

Der lexikographische Prozess im deutschen Wiktionary

1. Einführung

Wörterbücher werden traditionell von einer Wörterbuchredaktion erstellt. Dem gegenüber steht das *gemeinschaftliche Erarbeiten* der Wörterbuchinhalte, das ein völlig anderes Paradigma der Lexikographie darstellt. Die Wörterbucherstellung liegt hier nicht länger in den Händen einer Redaktion, sondern wird von einer meist großen Zahl freiwilliger Autoren übernommen. Das gemeinschaftliche Vorgehen wurde in der (meta-)lexikographischen Fachliteratur als *kollaborative Erstellung* (Storrer/Freese, 1996) und *bottom-up lexicography* (Carr, 1997) eingeführt, ist aber bis heute wenig erforscht. Insbesondere stellt sich für die Wörterbuchforschung die Frage, wie sich die freiwilligen Autoren organisieren und wie bei der Erarbeitung der Wörterbuchinhalte vorgegangen wird.

Im vorliegenden Beitrag analysieren wir das gemeinschaftlich erstellte Online-Wörterbuch *Wiktionary* hinsichtlich dieser Fragestellung, um insbesondere herauszuarbeiten, inwieweit sich der in Wiktionary zugrundeliegende lexikographische Prozess mit den Prozessen redaktionell erstellter Wörterbücher deckt bzw. sich davon unterscheidet. Unsere Auswertung bezieht sich auf das deutsche Wiktionary, das mit ca. 227.000 Artikelseiten¹ das derzeit größte deutschsprachige, gemeinschaftlich erarbeitete Wörterbuch darstellt und daher sowohl qualitative als auch quantitative Untersuchungen in dieser Forschungsfrage ermöglicht. Auf den Aufbau und die in Wiktionary kodierten Angaben werden wir im Rahmen dieses Beitrags nur kurz eingehen – für eine umfassende Darstellung verweisen wir auf Meyer/Gurevych (2012a).

In der Fachliteratur wird der lexikographische Prozess als Abfolge mehrerer Phasen beschrieben. Auf eine Vorbereitungs- oder Planungsphase folgen meist mehrere Phasen, in denen die Wörterbuchinhalte erarbeitet und für die Publikation aufbereitet werden. Wiegand et al. (2010, S. 8ff.) unterscheiden die folgenden fünf Phasen:

- Vorbereitungsphase
- Phase der Datenbeschaffung
- Phase der Datenaufbereitung
- Phase der Datenauswertung
- Phase der Satz- und Druckvorbereitung

In den nachstehenden Kapiteln werden wir diese Phasen mit redaktionell erstellten Wörterbüchern vergleichen und als Ergebnis dieser Arbeit ein neues Modell zur Beschreibung des lexikographischen Prozesses im gemeinschaftlich erstellten Wiktionary vorschlagen.²

¹ Die im Beitrag angegebenen Statistiken beziehen sich auf Wiktionary-Daten vom 5.7.2012.

² Es existieren zahlreiche andere Beschreibungen für lexikographische Prozesse, die auf bestimmte Wörterbuchttypen zugeschnitten sind oder auf den Computereinsatz umfassender eingehen (vgl. Müller-Spitzer, 2004; Klosa, 2013). Wir wählen die fünf Phasen von Wiegand et al. (2010), da diese für eine vergleichende Studie ausreichend allgemein formuliert und zudem in der Wörterbuchforschung hinreichend bekannt sind. Die zur Beschreibung der einzelnen Teilaufgaben nötige Terminologie werden wir jeweils kurz einführen und verweisen in Detailfragen auf Wiegand et al. (2010).

2. Vorbereitungsphase

In der Vorbereitungsphase gilt es unter anderem die Wörterbuchkonzeption auszuarbeiten und damit Wörterbuchtyp und Wörterbuchgegenstand festzulegen, den Lemmabestand zu planen und den Benutzerbezug zu definieren. In Wiktionary finden sich Informationen zur Wörterbuchkonzeption im sogenannten Autorenportal, einem separaten (nicht-lexikographischen) Bereich, der von den beitragenden Autoren zur Organisation und Planung des Wörterbuchs verwendet wird. Zum Wörterbuchtyp und -gegenstand heißt es hier:

Das Wikiwörterbuch ist das deutschsprachige Wiktionary: ein frei verfügbares, mehrsprachiges Wörterbuch für den Wortschatz aller Sprachen.³

[...] das Wiktionary wird nicht von einer festen, bezahlten Redaktion geschaffen, sondern ist vielmehr das Werk freiwilliger Autoren.⁴

Bei Wiktionary handelt es sich folglich um ein allgemeines, mehrsprachiges Wörterbuch, dessen Gegenstandsbereich sämtliche Einzelsprachen adressiert. Wie schon eingangs erwähnt, werden die Inhalte gemeinschaftlich statt redaktionell erarbeitet. Basierend auf den weiteren Beschreibungen im Autorenportal, kann Wiktionary als polyinformatives (mehrere sprachliche Eigenschaften beschreibend), polysektives (ohne Einschränkung auf eine bestimmte Klasse von Lemmazeichen) und polyakzessives (über mehrere Suchpfadtypen zugreifbare Einträge) Wörterbuch beschrieben werden. Neben der Suchmöglichkeit sind die Wörterbuchartikel über einen alphabetischen Index, nach Sachgruppen geordnet (bspw. Artikel aus dem Bereich Mathematik) und onomasiologisch mittels Synonym-, Hyponym- oder Hyperonymverweisen zugreifbar. Es finden sich hauptsächlich Angaben zur gegenwartsbezogenen „Standardsprache“, zum Teil sind aber auch Dialektausdrücke und Fachvokabular sowie Artikel zu veralteten Ausdrücken in Wiktionary enthalten. Um die Nachprüfbarkeit der in den Wörterbuchartikeln gemachten Angaben zu gewährleisten, sollen Belege angegeben werden:

Das Wiktionary ist [...] kein Forum zur Streuung von Wortschöpfungen oder von Wörtern, die in der Standard- oder Umgangssprache sehr selten benutzt werden. Dazu gehören Wörter, zu denen es keine nachprüfbaren, verständlichen Belege gibt. Eine Person, die das Wikiwörterbuch benutzt, sollte sich auch darauf verlassen können, dass die hier gefundenen Wörter wirklich geläufig sind.⁵

Diesen Aspekt werden wir in den Kapiteln 3 und 4 noch näher untersuchen. Wiktionary ist ein Ausbauwörterbuch; die Wörterbuchartikel sind direkt nach ihrer Erstellung online abrufbar und können zu jedem Zeitpunkt überarbeitet werden. Auf diesen Umstand werden wir in den Kapiteln 5 und 6 eingehen. Während sich Wiktionary in all diesen Bereichen anhand bestehender Wörterbuchtypologien einordnen lässt, bleibt der Benutzerbezug undefiniert. Es finden sich keine Angaben zum Adressatenkreis (Laien, Experten, Sprachenlerner, etc.) und zur Wörterbuchfunktion. Die von Wiktionary verwendete mehrsprachige Struktur schließt dabei weder die Rezeptions-, die Produktions-, die Hinübersetzungs- noch die Herübersetzungsfunktion aus.

Die Wörterbuchkonzeption wird üblicherweise durch ein lexikographisches Instruktionsbuch ergänzt, in dem sich praxisbezogene Informationen zu den einzelnen Phasen des lexikographischen Prozesses finden. Derartige Informationen sind in Wiktionary ebenfalls über das Autorenportal zugänglich. Neben einer allgemeinen Einführung für neue Autoren und technischer

³ <https://de.wiktionary.org/w/index.php?oldid=2431424> (Version vom 15.6.2012)

⁴ <https://de.wiktionary.org/w/index.php?oldid=2708038> (Version vom 21.10.2012)

⁵ <https://de.wiktionary.org/w/index.php?oldid=2175236> (Version vom 12.4.2012)

Dokumentation zur eingesetzten Software finden sich hier überwiegend Instruktionen zur Ausarbeitung der Wörterbuchartikel sowie zu einzelnen Artikelpositionen. Das Mikrostrukturprogramm wird neben den textuellen Erläuterungen durch eine Auswahl an Beispielartikeln und insbesondere durch verschiedene, wortartenspezifische Formatvorlagen definiert, welche die Grundlage für jeden neu erstellten Artikel bilden.

Während das Mikrostrukturprogramm in redaktionell erstellten Wörterbüchern üblicherweise im Rahmen der Vorbereitungsphase definiert wird, können in Wiktionary prinzipiell jederzeit Änderungen am Mikrostrukturprogramm erfolgen. Am Beispiel der Formatvorlage für Substantive haben wir untersucht, wie häufig diese Möglichkeit genutzt wird und welche Änderungen vorgenommen werden. Die erste Version einer Formatvorlage⁶ aus dem Jahr 2004 war noch nicht wortartenspezifisch. Vorgesehen waren die Artikelpositionen Sprache, Stichwortangabe, Wortart, Wortformen, Aussprache, Herkunft, Bedeutungen, Synonyme, Gegenteile, Oberbegriffe, Unterbegriffe, Beispiele, Redewendungen, Referenzen und weiterführende Informationen, abgeleitete Begriffe und Übersetzungen (in dieser Reihenfolge). Die einzelnen Artikelpositionen waren mit Beispielangaben zum Stichwort „Kraftfahrzeug“ und zum Teil mit kurzen Erläuterungstexten illustriert, bspw. „Automobil {bedeutungsgleiche oder bedeutungsverwandte Wörter}“ für die Synonymposition.

Gut ein Jahr später hatte sich das Mikrostrukturprogramm bereits stark verändert.⁷ Für verschiedene Wortarten wurden eigenständige Formatvorlagen angelegt, um bspw. auf Unterschiede in der Wortformangabe zwischen Substantiven und Verben eingehen zu können. Darüber hinaus finden sich die neu hinzugekommenen Artikelpositionen Silbentrennung, Abkürzungen, charakteristische Wortkombinationen, Dialektausdrücke und ähnliche Wörter. Die Artikelpositionen zu Übersetzungen und abgeleiteten Begriffen wurden vor die Referenzen verschoben. Zudem wurde die formatierte Darstellung des Wörterbuchartikels überarbeitet, sodass bspw. Übersetzungen nun als zweispaltige, farbig unterlegte Tabelle erscheinen statt als einspaltige Aufzählung.

Vergleicht man nun das Mikrostrukturprogramm von 2005 mit einer Version⁸ aus dem Jahr 2012, so finden sich vergleichsweise wenig Änderungen: Die Artikelposition Silbentrennung wurde in Worttrennung umbenannt und aus Gegenworten wurde die Position Gegenwörter. Die zuvor lemmaspezifisch angegebenen Ober- und Unterbegriffe sind nun lesartenbezogen anzugeben. Darüber hinaus wurden die Referenzen mit bereits vorbelegten Quellenangaben versehen (siehe dazu Kap. 3).

Unsere Analyse zeigt, dass in Wiktionary eine Vorbereitungsphase erkennbar ist, in der die Wörterbuchkonzeption und die lexikographischen Instruktionen definiert werden und insbesondere das Mikrostrukturprogramm intensiv ausgearbeitet wird. Nachträgliche Änderungen sind zwar zu jedem Zeitpunkt möglich, treten jedoch nach Abschluss der Vorbereitungsphase deutlich seltener auf. Dabei gilt es zu bedenken, dass Änderungen in den lexikographischen Instruktionen, insbesondere am Mikrostrukturprogramm, von den freiwilligen Autoren gemeinschaftlich in allen bereits angelegten Wörterbuchartikeln umgesetzt werden müssen. Daher wird jede Änderung zunächst unter den Autoren diskutiert und erst dann in die Seiten zum Mikrostrukturprogramm eingepflegt. Diese sind überdies derzeit nur von Administratoren veränderbar.

⁶ <http://de.wiktionary.org/w/index.php?oldid=13719> (Version vom 5.5.2004)

⁷ <http://de.wiktionary.org/w/index.php?oldid=99282> (Version vom 29.8.2005)

⁸ <http://de.wiktionary.org/w/index.php?oldid=2172258> (Version vom 9.4.2012)

3. Phase der Datenbeschaffung

Die wesentliche Aufgabe in der Phase der Datenbeschaffung ist der Aufbau einer Wörterbuchbasis, d.h. die Auswahl der primären, sekundären und tertiären Quellen, aus denen dann Belege für die lexikographischen Angaben gewonnen werden können. Die Menge der primären Quellen bildet dabei das lexikographische Korpus, bestehend aus Texten aus natürlichen Kommunikationssituationen, wie Zeitungsartikel, Belletristik, Fachliteratur etc. Nicht selten wird dabei auf ein ausgewogenes Verhältnis verschiedener Textsorten geachtet oder es werden nur Texte aus einer bestimmten Entstehungszeit ausgewählt. Die sekundären und tertiären Quellen bestehen aus anderen Wörterbüchern und weiterführenden sprachlichen Materialien (z.B. Grammatiken).

Bei redaktionell erstellten Wörterbüchern wird die Wörterbuchbasis als einheitlicher Quellenkatalog fest definiert. Wiktionary gibt dagegen keinen solchen einheitlichen Quellenkatalog vor, sondern überlässt die Auswahl geeigneter Quellen den Autoren. Eine Literaturliste⁹ mit ca. 750 häufig verwendeten Quellen wird von den Autoren gepflegt und dient als Hilfestellung zur Quellenangabe in den Wörterbuchartikeln. Es können jedoch jederzeit weitere, nicht in der Liste verzeichnete Quellen in einem Wörterbuchartikel angegeben werden. Die Literaturliste enthält zum größten Teil sekundäre und tertiäre Quellen, deren Nutzung bereits in den lexikographischen Instruktionen explizit angegeben ist:

Wie eingangs erwähnt brauchen wir Referenzen, um den Lesern und anderen Autoren die Gewissheit zu geben, dass die angegebenen Bedeutungen oder gar das Wort selbst nicht der Fantasie eines Einzelnen entsprungen sind und nur Bestandteil seines Idiolekts sind, sondern eine weitere Verbreitung haben oder hatten. Dies zeigt sich daran, dass schon andere Wörterbücher entsprechende Angaben enthalten, auf die man sich stützen kann, oder dass das Wort (in einer bestimmten Bedeutung) vermehrt in zitierbaren Sätzen gebraucht wird.¹⁰

In den Wörterbuchartikeln wird zwischen Referenzen und Einzelnachweisen unterschieden. *Referenzen* beziehen sich auf den Wörterbuchartikel als Ganzes und sollen das Wort und die angegebenen Bedeutungen belegen. *Einzelnachweise* sind direkt einer Artikelposition zugeordnet und werden vorwiegend als Beleg für Herkunftsangaben, Beispielsätze oder sonstige Angaben zur Wortverwendung eingesetzt.

Um die verwendeten Quellen analysieren zu können, haben wir alle 96.339 Wörterbuchartikel zu deutschen Stichwörtern¹¹ quantitativ ausgewertet. Es zeigte sich, dass 58% der Artikel mit Referenzen und 23% mit Einzelnachweisen versehen sind. Insgesamt fanden sich 241.207 Referenzen und 44.949 Einzelnachweise, was durchschnittlich 2,5 Referenzen und 0,47 Einzelnachweisen pro Artikel entspricht. Die angegebenen Referenzen und Einzelnachweise beziehen sich auf bis zu 29.000 eigenständige Quellen.¹² Tabelle 1 zeigt die 20 am häufigsten verwendeten Quellen und wie oft diese als Quellenangabe in Referenzen oder Einzelnachweisen verwendet werden. Die ersten sechs Quellen sind wenig überraschend, da ebendiese be-

⁹ <http://de.wiktionary.org/wiki/Wiktionary:Literaturliste> und Unterseiten (letzter Zugriff: 22.11.2012)

¹⁰ <http://de.wiktionary.org/w/index.php?oldid=1982189> (Version vom 2.10.2011)

¹¹ Unter den in Kapitel 1 erwähnten 227.000 Artikelseiten finden sich Wörterbuchartikel zu Stichwörtern unterschiedlicher Sprachen, bspw. der Artikel zum deutschen Stichwort „Gemüse“ und der Artikel zum englischen Stichwort „vegetable“. Beide Artikel sind (innerhalb des deutschen Wiktionarys) in deutscher Sprache beschrieben; siehe dazu auch Meyer/Gurevych (2012a, S. 263f.).

¹² Da die Quellenangaben in Wiktionary sehr heterogen sind, kann auf eine bestimmte Quelle in unterschiedlicher Weise verwiesen werden (bspw. mittels verschiedener URLs). Die Zahl der eigenständigen Quellen ist daher als obere Schranke zu verstehen.

reits durch die Formatvorlage als Referenz vorgegeben sind und sich damit automatisch in jedem neu erstellten Artikel finden. Dadurch bleibt unklar, ob diese Quellen tatsächlich in den Schreibprozess einfließen oder ungeprüft im Wörterbuchartikel verbleiben. Anders verhält es sich mit den übrigen Quellen: Diese müssen explizit durch die Autoren angegeben werden. Bei Online-Ressourcen wird die URL als Hyperlink angegeben, und bei Druckwerken findet sich üblicherweise ein Verweis auf konkrete Seitenzahlen oder logische Abschnitte, was darauf schließen lässt, dass diese Quellen tatsächlich von den Autoren eingesehen wurden und somit in den Schreibprozess eingeflossen sind.

1) Wikipedia	40.154	11) Duden: Rechtschreibung	1.514
2) Wortschatz-Lexikon (Univ. Leipzig)	38.515	12) Metzler Lexikon Sprache	1.005
3) canoo.net	37.579	13) Pfeifer: Etymologisches Wörterbuch	797
4) DWDS	33.596	14) Duden: Herkunftswörterbuch	751
5) TheFreeDictionary	12.741	15) Bußmann: Lexikon der Sprachwiss.	741
6) Deutsches Wörterbuch	10.173	16) Burkart: Vornamen	710
7) Duden online	6.693	17) OWID	698
8) Duden Universalwörterbuch	2.417	18) Duden: Vornamenlexikon	636
9) Duden: Großes Fremdwörterbuch	2.402	19) Brockhaus	532
10) Kluge: Etymologisches Wörterbuch	1.706	20) Herder	453

Tab. 1: Übersicht der 20 am häufigsten verwendeten Quellen in Wiktionary nebst Anzahl ihrer Verwendung als Referenz oder Einzelnachweis

Betrachtet man die Gesamtliste aller verwendeten Quellen, fällt erneut ein klarer Schwerpunkt auf sekundären und tertiären Quellen auf. Wie das Zitat zu Beginn dieses Kapitels vermuten lässt, stützen sich die Wiktionary-Autoren insbesondere beim Belegen der angegebenen Bedeutungen auf bereits existierende Wörterbücher, wie canoo.net, das Digitale Wörterbuch der Deutschen Sprache (DWDS), das Deutsche Wörterbuch oder Duden online. Primäre Quellen werden dagegen meist nur in Einzelnachweisen zu Beispielsätzen angegeben. Hier finden sich vorwiegend Digitalisate und online abrufbare Quellen, z.B. aus Online-Zeitungen und -Zeitschriften, digitalen Buchprojekten wie Google Books oder Projekt Gutenberg, sowie zu einem kleineren Teil Fachliteratur bspw. zu Rechtswissenschaften, Medizin oder Betriebswirtschaft.

4. Phase der Datenaufbereitung

Die Phase der Datenaufbereitung zielt auf die Erstellung der lexikographischen Zwischendatei ab, also einer geordneten Menge aufbereiteter Daten aus dem lexikographischen Korpus. Die lexikographische Zwischendatei bildet die Basis für das Formulieren der Wörterbuchartikel in der darauf folgenden Phase der Datenauswertung. Üblicherweise entsteht die lexikographische Zwischendatei durch systematisches Exzerpieren des lexikographischen Korpus, um die Belege für die späteren Wörterbuchartikel zu erfassen.

Im vorigen Kapitel haben wir dargelegt, dass in Wiktionary keine einheitliche Wörterbuchbasis und damit auch kein dezidiertes lexikographisches Korpus vorliegt. Dementsprechend findet auch keine systematische Exzerption wie in redaktionell erstellten Wörterbüchern statt und es existiert *de facto* keine lexikographische Zwischendatei. Die Phasen der Datenbeschaffung und Datenaufbereitung verschmelzen vielmehr mit der Phase der Datenauswertung, da die Wiktionary-Autoren unmittelbar beim Bearbeiten eines Wörterbuchartikels (bzw. einer bestimmten Artikelposition) neue Quellen beschaffen und neue Angaben auf Basis dieser Quel-

len belegen können. In redaktionell erstellten Wörterbüchern ist ein solches Vorgehen aufgrund der systematischen Datenaufbereitung die Ausnahme. Dies hat zum einen Konsistenzgründe – die primären Quellen sollen möglichst auf gleiche Weise exzerpiert werden – und zum anderen ökonomische Gründe, um nicht zu jedem Stichwort das gesamte lexikographische Korpus wieder neu prüfen und nach diesem Stichwort exzerpieren zu müssen, was die Bearbeitungszeit um ein Vielfaches verlängern würde. Im gemeinschaftlich erarbeiteten Wiktionary zeigt sich dagegen ein anderes Bild: Einerseits werden vorwiegend sekundäre und tertiäre Quellen als Belege verwendet, die bereits aufbereitet sind und keiner systematischen Exzerption bedürfen, andererseits verteilen sich die lexikographischen Aufgaben zur Erarbeitung eines Wörterbuchartikels auf viele Benutzer. Das deutsche Wiktionary hat etwa 36.000 registrierte Benutzer, wovon ca. 1.000 mehr als zehn Überarbeitungen vorgenommen haben. Die Suche nach geeigneten Belegen kann in einem solchen Wörterbuch daher ohne Weiteres dezentralisiert werden.

Ungeachtet der nicht erkennbaren systematischen Exzerption in Wiktionary, stellt sich die Frage, ob die lexikographischen Angaben quellenbasiert erarbeitet werden oder aus Sekundärquellen übernommen werden. Abbildung 1 zeigt den Wiktionary-Artikel zu „Betreuungsgeld“. In diesem Artikel finden sich drei Referenzen als Beleg für die angegebene Bedeutung, sowie ein Einzelnachweis zu einem Zeitungsartikel der Welt Online, aus dem der angegebene Beispielsatz entnommen wurde. Die übrigen Angaben zur Worttrennung, Aussprache und Herkunft sind nicht mit einer Quellenangabe versehen. Der referenzierte Wikipedia-Artikel liefert eine ausführliche, enzyklopädische Beschreibung zur Bedeutung und Entwicklung des Betreuungsgeldes. Das Digitale Wörterbuch der Deutschen Sprache gibt verschiedene Korpusbeispiele und signifikante Wortverbindungen an. Das Wortschatz-Lexikon der Universität Leipzig kodiert schließlich Angaben zur Häufigkeit, Morphologie und Grammatik, sowie weitere Korpusbeispiele. Keine dieser drei Quellen enthält Herkunfts- oder Ausspracheangaben. Die in Wiktionary angegebene Bedeutungsparaphrase „finanzielle, staatliche Unterstützung für Eltern, die ihr Kind zu Hause betreuen, anstatt es in eine Krippe zu schicken“, konnte in dieser Weise auch weder in einer der angegebenen Quellen noch in anderen häufig verwendeten Quellen, wie Duden online oder canoo.net nachgewiesen werden. Diese Beobachtung legt nahe, dass die gemachten Angaben, insbesondere die Bedeutungsparaphrase, auf Basis der Quellen und des eigenen Sprachgefühls der beteiligten Autoren erarbeitet wurden.

Weniger eindeutig ist die Sachlage in Artikeln wie z.B. „schneeweiß“: hier findet sich in Wiktionary die Bedeutungsparaphrase „weiß wie Schnee“ und bei Duden online „weiß wie [frisch gefallener] Schnee“.¹³ In diesen Fällen lässt sich selten eindeutig entscheiden, ob Angaben aus einer der Quellen übernommen (und ggf. leicht variiert wurden) oder unter Kenntnis der genannten Quellen erarbeitet wurden.

Wie wir bereits im vorigen Kapitel diskutiert haben, gilt dabei zu beachten, dass verhältnismäßig wenig primäre Quellen angegeben sind. Die fehlende Kenntnis von Korpusbeispielen zeigt sich auch in Hanks (2012) Analyse einiger Bedeutungsparaphrasen aus dem englischen Wiktionary. Dies legt eher die Erarbeitung der gemachten Angaben nahe statt einer Übernahme aus (korpusbasierten) Wörterbüchern.

¹³ <http://de.wiktionary.org/w/index.php?oldid=2504933> (Version vom 27.07.2012), <http://www.duden.de/rechtschreibung/schneeweiss> (letzter Zugriff: 11.01.2013)

Betreuungsgeld

Betreuungsgeld (Deutsch) [Bearbeiten]

Substantiv, n [Bearbeiten]

Worttrennung:
Be-treu-ungs-geld, Plural: Be-treu-ungs-gel-der

Aussprache:
IPA: [bəˈtʁuːŋs,ɡɛlt], Plural: [bəˈtʁuːŋs,ɡɛldɐ]
Hörbeispiele: —, Plural: —

Bedeutungen:
[1] finanzielle, staatliche Unterstützung für Eltern, die ihr Kind zu Hause **betreuen**, anstatt es in eine **Krippe** zu schicken

Herkunft:
Determinativkompositum aus den Substantiven *Betreuung* und *Geld* mit dem Fugenelement -s

Beispiele:
[1] „Merkel ist indes fest entschlossen, den seit Monaten schwelenden Streit über das *Betreuungsgeld* jetzt zu beenden: Noch in diesem Monat soll ein Gesetz auf den Weg gebracht werden, dass Mütter und Väter, die ihre Kleinstkinder nicht in eine Krippe geben, mit zunächst 100 Euro im Monat unterstützt werden.“^[1]

Übersetzungen [Bearbeiten]

Einklappen ▲

- Englisch: [1] [[]] → en
- Französisch: [1] [[]] → fr
- Italienisch: [1] [[]] → it
- Spanisch: [1] [[]] → es

? Referenzen und weiterführende Informationen:

- [1] Wikipedia-Artikel „Betreuungsgeld“
- [*] Digitales Wörterbuch der deutschen Sprache „Betreuungsgeld“
- [1] Uni Leipzig: Wortschatz-Lexikon „Betreuungsgeld“

Quellen:

1. ↑ „Das ist ein K.-o.-Schlag“. In: *Welt Online*. 15. Mai 2012, ISSN 0173-8437 (URL, abgerufen am 22. Mai 2012).

Abb. 1: Wiktionary-Artikel „Betreuungsgeld“¹⁴

Eng mit der Datenaufbereitung verbunden ist die äußere Selektion, also die Auswahl der Stichwörter, für die ein Wörterbuchartikel geschrieben werden soll. Bei redaktionell erstellten Wörterbüchern wird die Vorgehensweise zur Stichwortauswahl in der Regel durch das lexikographische Instruktionsbuch festgelegt. Die Bearbeitung der Stichwörter erfolgt dann gemäß einer bestimmten Ordnung, bspw. als Bearbeitungsteilwortschätze in alphabetischen Teilstrecken oder anhand lexikalisch-semantischer Kriterien aufgeteilt. In Wiktionary gibt es kaum Kriterien zur Stichwortauswahl. Nicht erwünscht sind im Allgemeinen Personen- und Markennamen, sowie Stichwörter, die nicht belegt wurden. Außerdem gibt es keine festgelegte Bearbeitungsreihenfolge. Die freiwilligen Autoren können vielmehr einzelne Stichwörter selbst auswählen und Artikel erstellen bzw. diese überarbeiten, ohne einem festen Organisationsplan zu folgen. Dadurch entstehen sowohl Einzelartikel als auch Artikelreihen aus bestimmten thematischen Bereichen, als fehlend oder wünschenswert gemeldete Artikel und Artikel aus alphabetischen Wortlisten (z.B. zum Grundwortschatz). Bestimmte Themenbereiche sind daher stärker ausgearbeitet als andere. Meyer/Gurevych (2010) beobachten im englischen Wiktionary bspw. eine größere Abdeckung von Bedeutungen aus Naturwissenschaften und Technik sowie im Bereich Sport.

¹⁴ <http://de.wiktionary.org/w/index.php?oldid=2690921> (Version vom 28.9.2012)

5. Phase der Datenauswertung

In der Phase der Datenauswertung werden die Wörterbuchartikel auf Basis der exzerpierten Belege verfasst. Die Lexikographen formulieren – je nach Organisationsplan – den gesamten Wörterbuchartikel oder eine bestimmte Artikelposition aus einem Bearbeitungsteilwortschatz (d.h. einer nach bestimmten Kriterien zusammengestellten Stichwortliste). Die offen angelegte Organisation in Wiktionary sieht keine derartigen Vorgaben vor. Ein Wörterbuchartikel kann zu jedem Zeitpunkt von einem beliebigen Autor begonnen werden. Auch muss weder der gesamte Artikel noch eine bestimmte Artikelposition komplett ausgefüllt werden. Das Prinzip des gemeinschaftlichen Erarbeitens basiert stattdessen auf der mehrfachen Überarbeitung des Artikels:

Diese Formatvorlage muss nicht beim Neuanlegen bzw. ersten Editieren komplett ausgefüllt werden. [...] Wer sich unsicher ist oder irgendetwas nicht weiß, lässt es einfach offen. Nachfolgende Bearbeiter können dies später auch noch eintragen.¹⁵

Die von Wiktionary verwendete Software speichert jede einzelne Überarbeitung der Artikel, sodass jederzeit nachvollziehbar bleibt, welcher Autor zu welchem Zeitpunkt eine bestimmte Änderung vorgenommen hat. In Anlehnung an Daxenberger/Gurevych (2012) unterscheiden wir dabei die Begriffe Änderungsoperation, Überarbeitung und Revision. Eine *Änderungsoperation* bezieht sich auf eine einzelne Artikelposition, die von einem Autor

- ergänzt (bspw. um eine Ausspracheangabe),
- abgewandelt (z.B. die Neuformulierung einer Bedeutungsparaphrase),
- gelöscht (z.B. das Entfernen eines Beispielsatzes),
- formatiert (typographische Änderungen, wie Fettdruck, Kursivschrift, etc.) oder
- auf eine frühere Version zurückgesetzt wird (bspw. bei Vandalismus).

Als *Überarbeitung* bezeichnen wir die Durchführung einer Menge von Änderungsoperationen, die ein bestimmter Autor zu einem bestimmten Zeitpunkt durchführt. Eine Überarbeitung kann mehrere Artikelpositionen betreffen und unterschiedliche Arten von Änderungsoperationen umfassen, z.B. das Hinzufügen einer Bedeutungsparaphrase zu einer neuen Lesart nebst Beispielsatz und Übersetzung (Ergänzung) bei gleichzeitigem Umformulieren einer bereits existierenden Bedeutungsparaphrase (Abwandlung). Das Ergebnis der Überarbeitung ist ein vollständiger Wörterbuchartikel zu einem bestimmten Zeitpunkt, was wir als eine *Revision* des Wörterbuchartikels bezeichnen. Für die Wörterbuchnutzer bleiben alle Revisionen eines Wiktionary-Artikels abrufbar, und es lassen sich die einzelnen Änderungsoperationen auch nach dem Zeitpunkt der Überarbeitung noch sichtbar machen.

In einer quantitativen Analyse haben wir alle 2,4 Millionen Revisionen der ca. 227.000 Wiktionary-Artikel ausgewertet. Es finden sich sehr viele Artikel mit nur einer einzigen Revision und wenige Artikel mit sehr vielen Revisionen. Diese Art der Verteilung wird als *Potenzgesetz* bezeichnet. Durchschnittlich bestehen für jeden Artikel etwa zehn Revisionen, davon sind drei bis vier durch eine Überarbeitung von registrierten Benutzern, weniger als eine durch anonyme, nicht-registrierte Benutzer und etwa sechs Revisionen durch die Überarbeitung von sogenannten Bots entstanden. Bots sind automatische Programme, die selbsttätig die Wörterbuchartikel durchlaufen und diese hinsichtlich einer bestimmten Aufgabe überarbeiten. Die Programmierung wird von den Wiktionary-Autoren gemeinschaftlich erstellt. Zumeist handelt

¹⁵ <http://de.wiktionary.org/w/index.php?oldid=2433266> (Version vom 17.6.2012)

es sich dabei um Formatierungsaufgaben (bspw. Formatieren in Fettdruck oder die Änderung der Reihenfolge der angegebenen Wortformen) sowie um Aufgaben, die die Konsistenz der Artikel sicherstellen sollen (z.B. nach einer Änderung der Formatvorlagen). Im Regelfall werden von Bots keine linguistischen Angaben ergänzt, weshalb wir uns im Folgenden auf die 838.000 Revisionen von registrierten Benutzern beschränken. Auch sind die von Bots durchgeführten Änderungen meist sehr fein granuliert, weswegen die Zahl der Überarbeitungen durch Bots nicht mit der von menschlichen Benutzern verglichen werden kann.

Etwa 75% der Revisionen wurden von den Autoren mit einem Kommentar versehen, der die Überarbeitung beschreibt, aber nicht Teil des eigentlichen Wörterbuchartikels ist. Nachfolgende Autoren können so besser nachvollziehen, welche Änderungen vorgenommen wurden und warum. Der Wiktionary-Artikel „Haus“ wurde mit 415 Revisionen am häufigsten überarbeitet. Es folgen die Artikel zum Substantiv „Wasser“ mit 338 Revisionen und zum Verb „gehen“ mit 303 Revisionen. Über 50% der 100 am häufigsten geänderten Artikel lassen sich dem Grundwortschatz zurechnen, was ein Vergleich mit der „GUT1 500“-Wortliste¹⁶ zeigt. Die zweite große Gruppe bilden Begriffe der Umgangssprache, insbesondere vulgäre Ausdrücke. Ein Artikel weist im Durchschnitt zwei bis drei unterschiedliche Autoren auf (wiederum einem Potenzgesetz folgend). Die Liste der Artikel mit den meisten Autoren deckt sich im Wesentlichen mit der Liste der häufig überarbeiteten Artikel, d.h. je häufiger ein Artikel überarbeitet wird, desto mehr Autoren wirken im Allgemeinen an der Erstellung mit. Die meisten Autoren finden sich für „gehen“ (77 Autoren), „Wasser“ und „Haus“ (jeweils 73 Autoren). Es ist wenig überraschend, dass ältere Artikel im Regelfall mehr Autoren aufweisen, da diese länger im Wörterbuch kodiert sind und die Wahrscheinlichkeit für eine häufige Überarbeitung höher als bei kürzlich angelegten Artikeln ist.

Wir bezeichnen denjenigen Autor mit den meisten Überarbeitungen eines Artikels als *Hauptautor*. Durchschnittlich nehmen Hauptautoren etwa 16% aller Überarbeitungen eines Wörterbuchartikels vor. Die Spanne reicht dabei von Artikeln, die komplett von einem Autor bearbeitet wurden, zu Artikeln, in denen der Hauptautor mit anderen Autoren nahezu gleichauf ist, ihm also keine besondere Rolle zukommt. Unsere Analyse der Hauptautorschaft zeigt, dass die Autoren in Wiktionary verschiedene Aufgaben wahrnehmen. Wir betrachten exemplarisch die Autoren #9200 und #720: Autor #9200 ist Hauptautor von 34 Artikeln, in welchen durchschnittlich 38% aller Überarbeitungen durch #9200 vorgenommen wurden. Insgesamt hat #9200 etwa 62% der Überarbeitungen in der Eigenschaft als Hauptautor vorgenommen, u.a. an den Artikeln „Privatrecht“, „Sozialrecht“, „Gleichheitsrecht“ (jeweils über 90% aller Überarbeitungen), „Sozialhilfe“ (36%) und „Arbeitslosengeld“ (9%). Bei Autor #9200 handelt es sich folglich um einen Autor mit Schwerpunktthema Rechts- und Sozialwissenschaften. Autor #720 ist Hauptautor von 8.891 Artikeln. Allerdings wurden in vielen Artikeln nur ein bis zwei Änderungen vorgenommen. Von den insgesamt 43.750 Überarbeitungen wurden nur 20% in der Eigenschaft als Hauptautor durchgeführt. Die Analyse dieser Artikel zeigt, dass #720 keinen thematischen Schwerpunkt verfolgt, sondern vorwiegend Übersetzungsangaben ergänzt und überarbeitet. In Wiktionary finden sich demnach sowohl Autoren, die ganze Artikel aus einem bestimmten Bearbeitungsteilwortschatz bearbeiten, als auch Autoren, die auf bestimmte Artikelpositionen spezialisiert sind.

Betrachtet man auch Artikel mit einem dominanten zweiten Hauptautor, d.h. Artikel, in denen zwei Autoren einen Großteil der Überarbeitungen vorgenommen haben, können die Zusammenarbeit und das Konfliktpotential bei der Artikelerarbeitung näher untersucht werden. Dies

¹⁶ www.gut1.de/grundwortschatz/grundwortschatz_500.html (letzter Zugriff: 24. 5.2012)

gilt insbesondere dann, wenn der zweite Hauptautor eine ähnlich hohe Zahl an Überarbeitungen vorgenommen hat wie der erste Hauptautor, was ein Indikator für gelungene Kooperation, aber auch für einen Konflikt sein kann. Im letztgenannten Fall wäre zu erwarten, dass die Autoren die vom jeweils anderen Autor beigetragenen Informationen wechselseitig überschreiben. In Wikipedia treten derartige Konflikte von Zeit zu Zeit auf; sie werden meist als *edit war* bezeichnet (vgl. Viégas et al. 2004). In einer qualitativen Auswertung von 20 Artikeln mit einem dominanten ersten und zweiten Hauptautor zeigte sich, dass die Hauptautoren entweder gemeinschaftlich-kooperativ oder zeitlich versetzt arbeiten. Der Artikel „backen“ wurde bspw. vom ersten Hauptautor vorwiegend ab 2010 bearbeitet, während der zweite Hauptautor im Jahr 2006 Überarbeitungen durchgeführt hat. Im Artikel „Evolution“ überschneiden sich die Bearbeitungszeiträume der beiden Hauptautoren, ohne dass Konflikte erkennbar wären. Inwieweit dies auf alle Autoren und Artikel – über die Hauptautoren hinaus – übertragbar ist, bleibt im Rahmen zukünftiger Studien zu klären.

Neben der Überarbeitung eines Artikels selbst steht zu jedem Artikel eine separate Diskussionsseite zur Verfügung. Auf dieser Seite finden sich keine lexikographischen Angaben, sondern es besteht die Möglichkeit, den Artikel als solchen zu diskutieren, Überarbeitungen zu planen und abzustimmen sowie bereits durchgeführte Überarbeitungen zu reflektieren. Auch anhand der Diskussionsseiten lassen sich die Vorgehensweise der Wiktionary-Autoren beobachten und insbesondere deren Zusammenarbeit und die auftretenden Konflikte analysieren. Auf der Diskussionsseite zum Artikel „Kreuzung“ findet sich bspw. folgender Eintrag:

- Entschuldigt meine heftige Wortwahl, aber ich weiß nicht, was hier für „Fachleute“ revertieren und sichten!!
- ein Ort, wo sich 2 Straßen treffen, nennt man „Straßenknick“ oder „Straßenecke“, im einfachsten Fall einfach nur Straße, wenn eine gerade verlaufende Straße von der a-Straße zur b-Straße wird.
 - ein Ort, wo sich 3 Straßen treffen, wird Straßengabel oder auch Abzweigung genannt
 - ein Ort, wo sich 4 oder mehr Straßen treffen, wird Kreuzung genannt. Das ist sinngleich mit der Aussage: ein Ort, wo sich 2 (oder mehr) Straßen kreuzen oder ein Ort, wo eine Straße eine zweite Straße quert.
- [...] Ich möchte jetzt hier keinen Editwar anzetteln, weshalb ich die stattgefundene Revertierung meiner Änderungen nicht erneut ändere und bitte, jemand mit Sachkunde und Verständnis möge sich der Angelegenheit annehmen.¹⁷

Dem Diskussionsbeitrag voraus geht eine Überarbeitung der Bedeutungsparaphrase von „Stelle, wo sich zwei oder mehrere Straßen treffen“ zu „Stelle, wo sich vier oder mehr Straßen treffen“, die kurze Zeit nach der Änderung wieder rückgängig gemacht wurde. Die weitere Diskussion zeigt zwei Interpretationen der Wortbedeutung, je nachdem ob eine Straße durch eine Kreuzung hindurchführt (zwei oder mehr Straßen treffen sich) oder diese in der Kreuzung endet (vier oder mehr Straßen treffen sich). Die Meinung des Überarbeiters kann sich im weiteren Verlauf nicht durchsetzen, und die Bedeutungsparaphrase verbleibt im Zustand vor der Überarbeitung. Derartige Beiträge zeigen, dass auf Diskussionsseiten die eigenen Sichtweisen bzw. die Interpretation von angeführten Quellen dargelegt und diskutiert werden und damit als Mittel zur Konfliktlösung eingesetzt werden.

6. Phase der Satz- und Druckvorbereitung

Die Phase der Satz- und Druckvorbereitung umfasst die Konzeption des endgültigen Layouts der Wörterbuchartikel und der begleitenden Komponenten (wie Um- und Binnentexte) für

¹⁷ <http://de.wiktionary.org/w/index.php?oldid=1992616> (Version vom 20.10.2011)

eine Print- oder Online-Publikation des Wörterbuchs. Die diesbezüglichen Abläufe in Wiktionary unterscheiden sich kaum von redaktionell erstellten Wörterbüchern: Die Wörterbuchinhalte in Wiktionary werden von den Autoren in einer Auszeichnungssprache (*Wikitext* oder *Wiki-Markup* genannt) kodiert und durch die MediaWiki-Software¹⁸ automatisch in fertig formatierte Wörterbuchartikel, Verzeichnisse, Begleittexte etc. umgesetzt. Analog dazu kodieren die Lexikographen in redaktionell erstellten Wörterbüchern die erarbeiteten Daten meist über ein Redaktionssystem, welches die Daten bspw. als XML speichert und eine automatische Formatierung (z.B. durch XSL-basierte Transformation) vornimmt.

Dabei gilt zu bemerken, dass in Wiktionary auch die Formatierung der Inhalte – wie bereits in den vorigen Kapiteln dargestellt – von den Autoren gemeinschaftlich erarbeitet und ausgehandelt wird. Dadurch kann sich als Ergebnis einer Diskussion die Darstellung der Wörterbuchinhalte prinzipiell grundlegend ändern. Die Benutzer haben folglich die Möglichkeit, nicht nur auf die Wörterbuchinhalte, sondern auch auf die Wörterbuchform einzuwirken. Wie die Analyse des Mikrostrukturprogramms in Kapitel 2 zeigt, spielt diese Option insbesondere zu Beginn der Wörterbucherstellung eine tragende Rolle.

Ein weiterer Unterschied zwischen Wiktionary und vielen redaktionell erstellten Wörterbüchern liegt in der fehlenden Trennung von Modellierung und Darstellungsform. Nicht selten erfolgt in redaktionell erstellten Wörterbüchern eine feingranulare Modellierung der kodierten Angaben, obgleich manche dieser Angaben im fertig formatierten Wörterbuch nicht mehr zu unterscheiden sind (vgl. Müller-Spitzer 2011, S. 49f.). Die in Wiktionary verwendete Auszeichnungssprache ist dagegen sehr nahe an der formatierten Darstellung. Die kodierten Bedeutungsparaphrasen werden bspw. bereits beim Schreiben des Wörterbuchartikels als Aufzählung ausgezeichnet. Zwar wird auch vermehrt eine Modellierung auf logischer Ebene durch die Nutzung von sogenannten Vorlagen (siehe dazu Meyer/Gurevych 2012) vorgenommen, jedoch verschwimmt die Phase der Satz- und Druckvorbereitung insgesamt stark mit der Phase der Datenauswertung, in welcher die Angaben kodiert werden.

7. Diskussion

Der Vergleich des lexikographischen Prozesses im deutschen Wiktionary mit dem redaktionell erstellten Wörterbüchern weist drei wesentliche Unterschiede auf:

1. Zum einen werden nachträgliche Änderungen von Wörterbuchartikeln von bisherigen Prozessbeschreibungen kaum abgedeckt. In redaktionell erstellten Wörterbüchern definiert die Wörterbuchkonzeption meist feste Abläufe zur inhaltlichen Prüfung und zum Korrekturlesen der erarbeiteten Wörterbuchartikel. Die Beteiligten sind dabei einer festen Rolle (Lexikographen, Hilfskräfte, Lektoren, etc.) zugeordnet. Aus ökonomischen Gründen sollten Korrekturarbeiten hier möglichst auf ein Minimum reduziert sein, d.h. die zur Korrektur vorgelegten Artikel sind idealerweise bereits komplett, konsistent und nachvollziehbar. Wie wir dagegen in Kapitel 5 gesehen haben, basiert das Schreiben von Wörterbuchartikeln in Wiktionary auf der mehrfachen Überarbeitung durch verschiedene Autoren. Die Wörterbuchartikel müssen vor einer Überarbeitung weder vollständig noch konsistent sein. Das nachträgliche Ändern eines Wörterbuchartikels und der Wörterbuchkonzeption wird damit zum zentralen Bestandteil des lexikographischen Prozesses.

¹⁸ www.mediawiki.org

2. Der zweite Unterschied liegt in der Trennung der einzelnen Phasen. Unsere Analyse hat gezeigt, dass die Phasen der Datenbeschaffung und Datenaufbereitung stark mit der Phase der Datenauswertung verschmelzen und dass ferner die Phase der Datenauswertung und die Phase der Satz- und Druckvorbereitung ebenfalls kaum zu unterscheiden sind.
3. Der dritte Unterschied liegt schließlich in der Diskussion der Autoren untereinander und mit den Benutzern. Sowohl zur Wörterbuchkonzeption als auch zu den Wörterbuchartikeln existieren gesonderte Diskussionsseiten, die zur Vorbereitung, Abwägung und Umsetzung von Änderungen genutzt werden können und die ein wichtiges Hilfsmittel zur Nachvollziehbarkeit von Änderungen darstellen.

Um den lexikographischen Prozess in Wiktionary besser beschreiben zu können, schlagen wir ein neues Prozessmodell vor, das in Abbildung 2 schematisch dargestellt ist. Unsere Beschreibung unterscheidet eine Vorbereitungs- und eine Bearbeitungsphase, in welchen eine Reihe von Schritten durchlaufen werden können. Die *Vorbereitungsphase* bildet – analog zu redaktionell erstellten Wörterbüchern – den Startpunkt eines Wörterbuchprojekts und adressiert die Ausarbeitung der Wörterbuchkonzeption sowie der lexikographischen Instruktionen. Im Anschluss daran (bzw. zum Teil auch parallel dazu) beginnt die *Bearbeitungsphase* mit der Erstellung der ersten Wörterbuchartikel. Der Artikelerstellungsschritt umfasst sowohl die Aufgaben zur Datenbeschaffung, -aufbereitung und -auswertung als auch zur Satz- und Druckvorbereitung eines Wörterbuchartikels. Diese einzelnen Teilaufgaben sind in Wiktionary kaum zu trennen und werden daher zusammengefasst. Wir unterscheiden davon allerdings die Artikelüberarbeitung. Während im Rahmen dieses Schritts ebenfalls Aufgaben zur Datenbeschaffung, -aufbereitung oder -auswertung durchgeführt werden, sind zusätzlich Korrektur- und Prüfaufgaben wahrzunehmen, die für die Artikelerstellung keine Rolle spielen. Die für die Artikelerstellung relevante Lemmaselektion wird von diesem Schritt dagegen nicht abgedeckt.

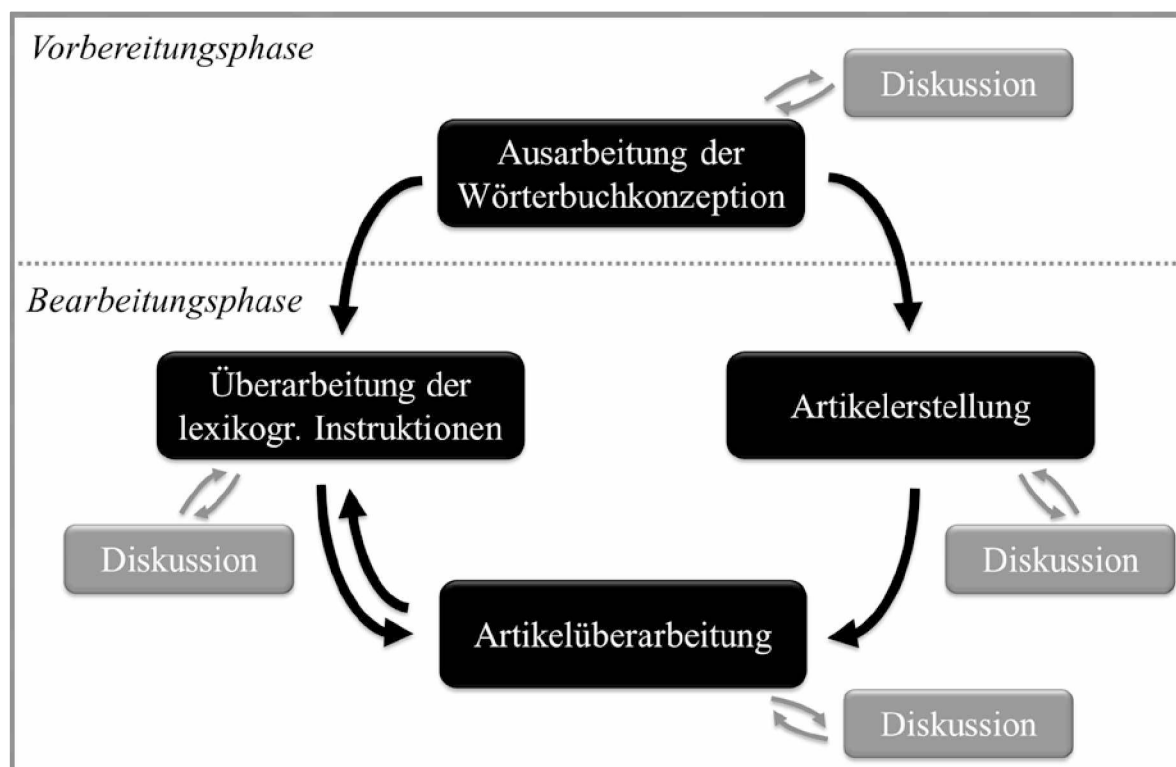


Abb. 2: Schema zur Beschreibung des lexikographischen Prozesses im deutschen Wiktionary.

Ergibt sich nach Abschluss der Vorbereitungsphase oder im Rahmen einer Artikelüberarbeitung ein Mangel oder eine wertvolle Ergänzung der lexikographischen Instruktionen, können diese ebenfalls noch nachträglich überarbeitet werden. Da derartige Änderungen zu Inkonsistenzen in den bestehenden Wörterbuchartikeln führen können, ergeben sich daraus wiederum Artikelüberarbeitungen, die von den Autoren (bzw. mittels automatischer Bot-Programme) über die Zeit hinweg vorgenommen werden müssen. Jeder der vier Schritte wird von einem Diskussionsprozess begleitet, in welchem die Arbeit organisiert oder die Vorgehensweise abgestimmt werden kann.

8. Zusammenfassung und Ausblick

Im vorliegenden Beitrag haben wir den lexikographischen Prozess im deutschen Wiktionary analysiert und mit den lexikographischen Prozessen von redaktionell erstellten Wörterbüchern verglichen. In Wiktionary ist eine Vorbereitungsphase erkennbar, in der die Wörterbuchkonzeption ausgearbeitet wird. Die lexikographischen Instruktionen lassen sich auch nach Abschluss dieser Vorbereitungsphase noch verändern, z.B. hinsichtlich des Mikrostrukturprogramms. Derartige Änderungen erwiesen sich nach Abschluss der Vorbereitungsphase als wenig umfangreich. Quellenangaben kommt in Wiktionary eine wesentliche Bedeutung zu, um die lexikographischen Angaben zu belegen. Es existiert jedoch kein einheitlicher Quellenkatalog; die Autoren können vielmehr beliebige Quellen auswählen. Ein deutlicher Schwerpunkt liegt dabei auf sekundären Quellen, d.h. die Wiktionary-Autoren stützen sich bei der Ausarbeitung der Wörterbuchartikel stark auf andere Wörterbuchprojekte. Ob die Angaben aus den Sekundärquellen übernommen oder auf Basis von Quellen erarbeitet werden, ließ sich nicht abschließend klären. Eine systematische Exzerption eines lexikographischen Korpus, wie bei redaktionell erstellten Wörterbüchern üblich, war in Wiktionary jedenfalls nicht zu erkennen.

Die Phasen zur Datenbeschaffung, -aufbereitung und -auswertung sind in Wiktionary nicht klar zu trennen. Stattdessen basiert der lexikographische Prozess auf der mehrfachen Überarbeitung der Wörterbuchartikel. Diese haben zwischen einem und 77 Autoren, durchschnittlich zwei bis drei. Häufig existiert ein sogenannter Hauptautor, welcher einen Großteil der Überarbeitungen vorgenommen hat. Soweit mehrere Hauptautoren am gleichen Artikel arbeiteten, konnten wir eher ein kooperatives oder zeitlich versetztes Arbeiten beobachten. Konflikte wurden eher auf Diskussionsseiten ausgetragen, die für alle Phasen des lexikographischen Prozesses zur Verfügung stehen.

Die häufige Überarbeitung der Wörterbuchartikel wird durch bestehende Beschreibungen des lexikographischen Prozesses kaum abgedeckt. Aus diesem Grund schlagen wir ein neues Prozessmodell zur Beschreibung der Vorgänge im gemeinschaftlich erstellten Wiktionary vor. In zukünftigen Arbeiten bleibt zu klären, inwieweit sich die lexikographischen Prozesse verschiedener gemeinschaftlich erstellter Wörterbücher ähneln und wie die Abläufe in redaktionell erstellten Wörterbüchern dadurch verbessert werden können.

9. Danksagung

Diese Arbeit wurde im Rahmen des Lichtenberg-Programms durch die VolkswagenStiftung gefördert (Geschäftszeichen I/82806). Wir danken Prof. Dr. Angelika Storrer für ihre hilfreichen Anmerkungen.

10. Literaturverzeichnis

- Carr, Michael (1997): Internet dictionaries and lexicography. In: *International Journal of Lexicography* 10, 3, S. 209-230.
- Daxenberger, Johannes/Gurevych, Iryna (2012): A corpus-based study of edit categories in featured and non-featured Wikipedia Articles. In: *Proceedings of the 24th International Conference on Computational Linguistics*, Mumbai, Indien, S. 711-726.
- Hanks, Patrick (2012): Word meaning and word use: corpus evidence and electronic lexicography. In: Granger, Sylviane/Paquot, Magali (Hg.): *Electronic lexicography*. Oxford: Oxford University Press, S. 57-82 [= Kapitel 4].
- Klosa, Annette (2013): The lexicographical process (with special focus on online dictionaries). In: Gouws, Rufus Hjalmar/Heid, Ulrich/Schweickard, Wolfgang/Wiegand, Herbert Ernst (Hg.): *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. (= *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK)* 5.4). Berlin/Boston, S. 517-524.
- Meyer, Christian M./Gurevych, Iryna (2010): How web communities analyze human language: Word senses in Wiktionary. In: *Proceedings of the Second Web Science Conference*. Raleigh, NC, USA.
- Meyer, Christian M./Gurevych, Iryna (2012): OntoWiktionary – Constructing an ontology from the collaborative online dictionary Wiktionary. In: Pazienza, Maria Teresa/Stellato, Armando (Hg.): *Semi-automatic Ontology Development: Processes and Resources*. Hershey, PA: IGI Global, S. 131-161 [= Kapitel 6].
- Meyer, Christian M./Gurevych, Iryna (2012a): Wiktionary: A new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography. In: Granger, Sylviane/Paquot, Magali (Hg.): *Electronic Lexicography*. Oxford: Oxford University Press, S. 259-291 [= Kapitel 13].
- Müller-Spitzer, Carolin (2004): Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexikographischen Prozessen. In: *Lexicographica* 19, S. 140-168.
- Müller-Spitzer, Carolin (2011): Der Aufbau einer maßgeschneiderten XML-basierten Modellierung für ein Wörterbuchnetz. In: Klosa, Annette/Müller-Spitzer, Carolin (Hg.): *Datenmodellierung für Internetwörterbücher*. 1. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“ (= OPAL – Online publizierte Arbeiten zur Linguistik 2/2011), Mannheim: Institut für Deutsche Sprache, S. 37-51.
- Storrer, Angelika/Freese, Katrin (1996): Wörterbücher im Internet. In: *Deutsche Sprache* 24, 2, S. 97-153.
- Viégas, Fernanda B./Wattenberg, Martin/Dave, Kushal (2004): Studying Cooperation and Conflict between Authors with History Flow Visualizations. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Wien, Österreich, S. 575-582.
- Wiegand, Herbert Ernst/Beißwenger, Michael/Gouws, Rufus H./Kammerer, Matthias/Storrer, Angelika/Wolski, Werner (Hg.) (2010): *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research*. Bd. 1 (= *Systematische Einführung / Systematic Introduction*, A-C). Berlin/New York: de Gruyter.
- Wiktionary (2002 et seq.), San Francisco: Wikimedia Foundation. www.wiktionary.org. (Stand: Februar 2015).

Rufus H. Gouws

Aspekte des lexikographischen Prozesses in Print- und Onlinewörterbüchern

1. Einleitung

Nach McArthur (1986) sind Wörterbücher „containers of knowledge“. Das gilt für Printwörterbücher, aber auch für elektronische Wörterbücher.¹ Aus beiden Wörterbuchtypen kann man häufig die gleiche Information erhalten. Obwohl der Inhalt nicht immer unterschiedlich ist, sehen diese „Containers of knowledge“ nicht ähnlich aus. Jedes Wörterbuch bleibt aber das Produkt eines lexikographischen Prozesses. Diese Prozesse sind für verschiedene Wörterbücher unterschiedlich. In jedem Prozess gibt es bestimmte Phasen, und häufig findet man dieselben Phasen bei der Erarbeitung von Print- und Internetwörterbüchern.

In diesem Beitrag werden vergleichend verschiedene Aspekte des lexikographischen Prozesses von Print- und Internetwörterbüchern besprochen, ohne dabei jedoch alle Aspekte des jeweiligen lexikographischen Prozesses zu behandeln. Die Aufmerksamkeit wird vor allem auf bestimmte lexikographische Strukturen gelenkt.

Ob man, wie Landau (1984), die Lexikographie als ein „Art or craft“ oder wie Wiegand als eine wissenschaftliche Praxis betrachtet, ist es so, dass diese Praxis immer gute Planung braucht, die in die Formulierung eines lexikographischen Prozesses mündet. Die Idee lexikographischer Planung ist nicht neu. Für sein Wörterbuch aus dem Jahre 1755 hatte Samuel Johnson schon seinen *Plan of a Dictionary of the English Language* (Johnson 1747). Wichtig für die Zuverlässigkeit eines Wörterbuchs ist die konsistente Darstellung und Bearbeitung der lexikographischen Daten. Lexikographische Konsistenz erhöht die Benutzerfreundlichkeit eines Wörterbuchs, die Auffindbarkeit der lexikographischen Daten und ermöglicht erfolgreiche Wörterbuchbenutzungshandlungen. Das soll das Ziel jedes Wörterbuchs und darum auch jedes lexikographischen Prozesses sein.

In Wiegand (1998) findet man eine ausführliche Diskussion verschiedener lexikographischer Prozesse, wobei Aspekte lexikographischer Prozesse ohne Computereinsatz, Aspekte computerunterstützter lexikographischer Prozesse und die Einordnung der computerlexikographischen Prozesse behandelt werden. In diesem Beitrag geht es nicht um computerunterstützte lexikographische Prozesse, sondern um lexikographische Prozesse, deren Produkte Internetwörterbücher sind, Wörterbücher, die nicht auf einem Printwörterbuch basieren, sondern unmittelbar für das Online-Medium konzipiert sind.

In Internetwörterbüchern sieht der lexikographische Prozess in bestimmten Hinsichten anders aus als in Printwörterbüchern. Einige dieser Unterschiede werden in diesem Beitrag genannt und besprochen.

Dabei ist es wichtig, den lexikographischen Prozess für Internetwörterbücher nicht als isoliertes oder theoriefreies Objekt zu betrachten. Ein Vergleich der verschiedenen Typen von lexikographischen Prozessen soll auch zu einem Vergleich verschiedener Typen von Lexikographien führen. Auch das wird dieser Beitrag leisten.

¹ Ich danke Ulrich Heid, Vera Hildenbrandt und Annette Klosa für Kommentare zu Sprache und Inhalt dieses Beitrags.

2. Eine allgemeine Lexikographietheorie

Es gibt verschiedene und auch unterschiedliche Meinungen über Lexikographietheorie. In diesem Beitrag geht es nicht um eine bestimmte Theorie oder eine bestimmte Version einer Theorie. Wichtig ist, dass eine erfolgreiche lexikographische Praxis eine Theorie braucht.

In der Entwicklung der theoretischen Lexikographie kann man verschiedene Phasen unterscheiden. Nach Gouws (2005) ist die erste Phase die sogenannte Zgusta-Phase – mit dem linguistischen Inhalt eines Wörterbuchs als Hauptforschungsbereich. Für die zweite oder ‘Wiegand-Phase’ bleibt der Inhalt wichtig, aber die Verpackung der Daten rückt in den Vordergrund, und darum stehen die Wörterbuchstrukturen im Mittelpunkt der Diskussion. Der anvisierte Benutzer wird immer wichtiger, und in der dritten Phase, der sogenannten ‘Bergenholtz/Tarp-Phase’, spielen Wörterbuchfunktionen die wichtigste Rolle. Die Schwerpunkte dieser drei Phasen sind alle wichtig für eine erfolgreiche Lexikographietheorie.

Bevor der lexikographische Prozess für Internetwörterbücher betrachtet wird, muss die derzeitige Lexikographietheorie fokussiert werden. Der Besprechung der verschiedenen lexikographischen Prozesse in Wiegand (1998) liegt die damalige Theorie zugrunde, die sich vor allem auf Printwörterbücher richtete, obwohl Wiegand (ebd.) auch den Prozess mit Computereinsatz erklärt. Für die Formulierung eines lexikographischen Prozesses für Internetwörterbücher brauchen wir heute aber eine allgemeine Lexikographietheorie, die auch Internetwörterbücher einbezieht. Leider existiert eine solche Theorie bislang noch nicht. Nach Zgusta (1971) ist die Lexikographie eine Tätigkeit, in der Tradition eine wichtige Rolle spielt, und eine dieser Traditionen ist, dass die Lexikographie sich sehr langsam an neue Umstände gewöhnt.

Zgusta (1971, S. 20) konstatiert aber auch: „The lexicographer's work is always creative, in a greater or a lesser degree, because he must always try to find new solutions to problems as yet unsolved.“ Die Metalexikographie steht daher heutzutage vor der großen Herausforderung, auch Internetwörterbücher berücksichtigen zu müssen.

Eine wichtige Entscheidung in dieser Hinsicht ist, ob man eine neue Theorie für die e-Lexikographie braucht und dann von zwei verschiedenen Theorien für Print- und für Internetwörterbücher ausgeht oder ob man nur eine allgemeine Theorie aufstellt, die für Print- und e-Lexikographie gilt. Meiner Meinung nach wird nur eine allgemeine Lexikographietheorie benötigt, die als Basis für die Formulierung aller Arten von lexikographischen Prozessen dienen kann.

Meine eigene Einstellung gegenüber der Lexikographietheorie ist die folgende: Mit den Theorien von Wiegand und Bergenholtz/Tarp gibt es zwei starke Kandidaten, die miteinander kombiniert werden sollten, da (in der Planung von Print- und Internetwörterbüchern) sowohl Wörterbuchfunktionen wie auch Wörterbuchstrukturen und ein Fokus auf den Wörterbuchinhalt sehr wichtig sind. Nach der Identifizierung der Benutzergruppe und der Bedürfnisse dieser Gruppe sollen die Funktionen eines bestimmten Wörterbuches festgelegt werden. Diese Funktionen bestimmen dann die jeweiligen Daten und Wörterbuchstrukturen. In einer allgemeinen Lexikographietheorie sollen Wörterbuchfunktionen also eine ganz wichtige Rolle spielen, aber als eine Unterabteilung der allgemeinen Theorie und nicht als eine selbständige Theorie. Wörterbuchfunktionen wie auch die Theorie der lexikographischen Funktionen sind medienneutral formuliert. Das ist ein erheblicher Vorteil. Die sogenannte allgemeine Lexikographietheorie von Wiegand ist vor allem eine Theorie für Printwörterbücher. Man braucht heutzutage eine

Neukonzeptionierung mit bestimmten Änderungen, um die Medienneutralität der Theorie zu gewährleisten. Bestimmte Unterabteilungen dieser Theorie beziehen sich dann primär auf Print-, andere Unterabteilungen primär auf Internetwörterbücher, und wieder andere Unterabteilungen dienen sowohl für Print- als auch für Internetwörterbücher.

Die Lexikographietheorie ist viel jünger als die lexikographische Praxis. Die Formulierung der früheren Theorie war oft eine Reaktion auf Aspekte der lexikographischen Praxis und ein Versuch, die Praxis zu beschreiben. Sie folgte der Praxis, statt ihr vorauszugehen. Eine – bislang fehlende – umfassende und allgemeine Theorie, die sich auf Internetlexikographie richtet, sollte nicht nur die aktuelle Internetwörterbuchpraxis beschreiben, sondern sie gestalten, indem sie neue Vorschläge macht, für neue und bessere Wörterbücher. Die Verbesserung der lexikographischen Praxis, auch der Praxis der Internetlexikographie, soll auf der Basis einer solchen Theorie erfolgen. In dieser Hinsicht ist es wichtig, auch eine theorie-basierte Neuformulierung des lexikographischen Prozesses zu haben, eines Prozesses, der den praktischen Lexikographen helfen kann, bessere Wörterbücher zu machen.

3. Der lexikographische Prozess

3.1 Printwörterbücher

Wenn man sich mit einem lexikographischen Prozess beschäftigt, handelt es sich entweder um die Formulierung eines theoretischen Modells, das für die praktische Anwendung entwickelt wird, oder um die praktische Anwendung eines solchen Modells, die aus aufeinanderfolgenden Anwendungshandlungen besteht. In diesem Beitrag geht es nur um die theoretische Formulierung des lexikographischen Prozesses, obwohl es sehr wichtig ist, dass ein solches Modell nie nur eine Menge abstrakter Vorschläge bleibt. Bestimmte Vorschläge dieses Beitrags müssen noch in der Praxis geprüft werden.

Nach Wiegand hat ein lexikographischer Prozess (ohne Computereinsatz) die folgende Einteilung:

- eine Vorbereitungsphase,
- eine Phase der Materialbeschaffung,
- eine Phase der Materialaufbereitung,
- eine Phase der Materialauswertung,
- eine Phase der Satz- und Druckvorbereitung.

Die Vorbereitungsphase, die für die Thematik dieses Beitrags wichtig ist, führt zu einem Wörterbuchprogramm, indem sie auf die Erarbeitung eines Wörterbuchplans abzielt, der aus einem Organisationsplan und einer Wörterbuchkonzeption besteht. Die Wörterbuchkonzeption stellt den eigentlichen lexikographischen Teil eines Wörterbuchplans dar. „Einzugehen ist in diesem Teil auf die Zwecke des Wörterbuchs, den Wörterbuchtyp, den Benutzerbezug, auf den Wörterbuchgegenstand, auf den Lemmabestand, auf die vorgesehenen Außentexte [...] und die Wörterbuchbasis.“ (Wiegand 1998, S. 137).

3.2 Internetwörterbücher

Mit mehr oder weniger Änderungen hat der lexikographische Prozess für elektronische Wörterbücher ähnliche Phasen wie derjenige von Printwörterbüchern. In diesem Beitrag werden nicht alle verschiedenen Phasen oder der ganze Prozess besprochen, sondern nur einige Aspekte einiger Unterabteilungen, ohne zu versuchen, ein zusammenhängendes Modell zu bieten.

3.2.1 Der Organisationsplan

Wie bei den lexikographischen Prozessen für Printwörterbücher enthält das Wörterbuchprogramm einen Organisationsplan und eine Wörterbuchkonzeption. Einer der Teilpläne des Organisationsplans ist ein Personalplan. Auch hier gibt es ein Herausgeberteam mit verschiedenen Fähigkeiten und Aufgaben.

Nach Wiegand (1998, S. 244) stammen die Kenntnisse, die erforderlich sind, um Handlungen ausführen zu können, im Bereich eines computerlexikographischen Prozesses überwiegend aus der Informatik und der Computerlinguistik. Er sagt: „Während im computerunterstützten lexikographischen Prozess neben den Lexikographen ein Informatiker benötigt wird, wird im computerlexikographischen Prozess neben Informatikern und Computerlinguisten auch ein Lexikograph benötigt bzw. jemand, der lexikographische und metalexikographische Kenntnisse besitzt.“ Die verschiedenen Kompetenzen werden also in den verschiedenen Prozessen unterschiedlich gewichtet.

Nach Müller-Spitzer (2003, S. 156) ist die Feststellung Wiegands für ihn der Grund „die beiden Arten von Prozessen dem Gegenstandsbereich verschiedener wissenschaftlicher Disziplinen zuzuordnen“: Nach Wiegand werden die beiden Arten von Prozessen von Wissenschaftlern in Gang gehalten, welche eine unterschiedliche akademische Ausbildung erfahren haben, und darum soll man die computerlexikographischen Prozesse nicht zum Gegenstandsbereich der Wörterbuchforschung rechnen. Das war eine Beurteilung des lexikographischen Prozesses aus der Sicht der Printwörterbuchmetalexikographie. In der Planung eines lexikographischen Prozesses für elektronische Wörterbücher müssen wir aber den Bereich der Wörterbuchforschung neu überdenken. Das ist eine wichtige und neue Aufgabe des lexikographischen Prozesses. Alle Aspekte eines lexikographischen Prozesses gehören zum Bereich der Wörterbuchforschung. Wir plädieren also für einen breiteren metalexikographischen Bereich, und das soll aus dem lexikographischen Prozess klar werden. Wir müssen neu nachdenken über die Wörterbuchforschung, über die Lexikographietheorie und die lexikographischen Prozesse. Das soll aber auch dazu führen, neu nachzudenken über Lexikographen und über die Teilnehmer an einem lexikographischen Prozess. Für Internetwörterbücher brauchen wir neue Typen von Mitarbeitern, die, obwohl sie eine andere akademische Ausbildung erfahren haben, in einem Wörterbuchprojekt als Lexikographen arbeiten. Also Lexikographen in einem weiter gefassten Sinn. Das Herausgeberteam eines Internetwörterbuchs muss anders aussehen als das Team eines Printwörterbuchs. In einem solchen Team finden wir Mitglieder mit einer anderen Ausbildung als die Lexikographen, z.B. als Informatiker, Lexikographen, die andere Aufgaben haben als die traditionelle Wörterbucharbeit, aber ihre Arbeit ist genauso wichtig für den spezifischen lexikographischen Prozess wie die Arbeit der Mitarbeiter mit einer lexikographischen Ausbildung.

Der Personalplan eines lexikographischen Prozesses für Internetwörterbücher verlangt nach Änderungen im Vergleich zu dem für Printwörterbücher. Wiegand (2012) plädiert für den Terminus *Lexico-engineer* statt *Lexikograph*. Der *-graph* in *Lexikograph* sei nämlich nicht relevant in einem lexikographischen Prozess für Internetwörterbücher. Gouws (2011, S. 25f.) behauptet, dass „a broadening in the expertise base of lexicographers has been established“ und dass „[w]e need a different species of role players, i.e. a new generation of lexicographers, and identifying these lexicographers must be the result of pondering on the question ‘Who is a lexicographer?’“. Wir brauchen eine neue Antwort auf diese Frage. Heutzutage ist ein Lexikograph wahrscheinlich nicht nur jemand mit einer traditionellen lexikographischen (= linguistischen) Ausbildung, sondern auch jemand mit einer spezifischen nicht-lexikographischen und vor allem nicht-linguistischen Ausbildung, z.B. als Informatiker. Diese weitere Ausbildung ist jedoch notwendig für den Erfolg eines lexikographischen Prozesses.

Ein neuer Teil des Organisationsplans ist eine Bestimmung der verschiedenen Kategorien der Mitarbeiter des Prozesses, aber auch das Aufstellen von Richtlinien für die spezifische akademische Ausbildung der verschiedenen Lexikographen. Das führt zu einer Erweiterung des Organisationsplans.

3.2.2 Die Wörterbuchkonzeption

Eine neue Phase der Wörterbuchkonzeption kann eine Entwurfsphase sein. Das ist eine frühe Phase im Prozess, in der Lexikographen Entscheidungen treffen über verschiedene Aspekte des Entwurfs des geplanten Wörterbuchs. Es handelt sich hier um Dinge wie Layout, Benutzerschnittstelle usw. Es gibt eine Menge elektronische Merkmale, die nicht für Printwörterbücher relevant sind, z.B. Verweise mit Hilfe von Links, sogenannte „expandable and collapsable indications“, Maus-sensitive Bereiche und Pop-up-Menüs mit lexikographischen Daten, die der Benutzer auf Wunsch öffnen kann, wenn er den Mauszeiger dorthin bewegt. Solche „Pop-up-Menüs“ können aus der Sicht eines Printwörterbuchs als *Infokästen* bezeichnet werden. Im lexikographischen Prozess eines Printwörterbuchs muss der Lexikograph entscheiden, ob er eingelagerte Binnentexte verwenden will. Die gleichen Daten, die in einem Printwörterbuch in diesen Texten erscheinen, können in einem elektronischen Wörterbuch in einem Pop-up-Menü erscheinen. In der Entwurfs- oder Einrichtungsphase muss der Lexikograph über viele neue Merkmale nachdenken, z.B. wo der Lexikograph eines Printwörterbuchs verschiedene Umtexte planen soll, kann der Lexikograph eines elektronischen Wörterbuchs eine Menge „outer features“ planen, die auch Wörterbuch-extern sein können. Nach Prinsloo et al. (2012, S. 295) handelt es sich hier um „devices for on-demand data provision“. Ein solcher Entwurf gibt jedem Benutzer die Gelegenheit, für sich zu entscheiden, welche und wie viele Daten er sehen will. Diese Phase ist eng verbunden mit der Implementierungsphase des Wörterbuchs – eine Phase, die sehr wichtig ist, aber nicht in diesem Beitrag besprochen wird.

Ein Teil der Entwurfsphase, der für Printwörterbücher nicht relevant ist, ist die Entscheidung, ob es sich bei einem bestimmten lexikographischen Prozess um ein Modell für ein einzelnes Wörterbuch handelt oder um ein Modell für eine Wörterbuchreihe auf der Grundlage einer gemeinsamen Databasis, d.h. eine datenbasisbasierte Wörterbuchreihe. Noch umfänglicher wird der Entscheidungsprozess, wenn sich der Prozess nicht nur auf Wörterbücher, sondern auch auf andere Nachschlagewerke richtet. Hier kann man den Terminus *Gesamtprozess* von Wiegand neu verwenden. Er deutet auf die verschiedenen Einzelprozesse, die für die verschiedenen Wörterbücher in einer datenbasisbasierten Wörterbuchreihe vorkommen und ge-

meinsam betrachtet werden müssen. In einer solchen Datenbasis findet man Daten, die für verschiedene Wörterbücher entnommen werden können – je nach dem Profil des anvisierten Benutzers.

So brauchen wir zum Beispiel ein neues Wörterbuch mit Afrikaans und Deutsch als Sprachenpaar. Ein Verlag hat derzeit wenig Interesse an einem solchen Wörterbuch, weil es unterschiedliche kleine Benutzergruppen gibt und kein einzelnes Wörterbuch die unterschiedlichen Bedürfnisse dieser Benutzergruppen befriedigen kann und weil es zu kostenintensiv ist, fünf verschiedene Wörterbücher zu produzieren – für Schüler, Studenten, Touristen, Wirtschaftler, allgemeine Benutzer. In jeder einzelnen Gruppe gibt es weitere unterschiedliche Nutzeranforderungen, z.B. solche, die sich auf Textproduktion und Textrezeption richten, die Bedürfnisse der Muttersprachler und der Nicht-Muttersprachler usw. In einer Datenbasis kann der Lexikograph Dateneinträge haben, die für jedes Bedürfnis und für jeden Benutzer markiert sind. Jeder Benutzer legt sein eigenes Profil fest, und nach seinen Ansprüchen wird Information aus der Datenbasis entnommen. Hier handelt es sich nicht mehr um eine typologiegetriebene Klassifikation von Wörterbüchern, sondern um eine merkmalsgetriebene Klassifikation, und diese Merkmale folgen aus dem Profil des individuellen Benutzers und seiner spezifischen Bedürfnisse. Der Gesamtprozess wird also nicht ein lexikographischer Prozess eines einzelnen typologisch klassifizierten Wörterbuchs, z.B. eines einsprachigen englischen Lernerwörterbuchs, sein, sondern ein Prozess einer Wörterbuchreihe mit einer Menge lexikographischer Merkmale. Diese Merkmale können sich in verschiedenen Kombinationen verbinden.

Die Planung einer solchen Datenbasis soll am Anfang des lexikographischen Prozesses vorgenommen werden – während der Entwurfsphase der Wörterbuchkonzeption. Eine solche datenbasisbasierte Wörterbuchreihe wird ggf. nicht nur Wörterbücher enthalten, sondern möglicherweise auch andere relevante Nachschlagewerke. Daraus kann sich dann ein von Müller-Spitzer (2003, S. 148) so genanntes Wortschatzinformationssystem entwickeln. Lexikographische Prinzipien werden in einem solchen System auch für die Planung anderer Nachschlagewerke verwendet, die auf derselben Datenbasis beruhen wie die Wörterbuchreihe. Es handelt sich hier nicht um eine Art lexikographischen Kolonialismus, bei dem die Lexikographietheorie andere Nachschlagewerke zu kolonisieren versucht, sondern eher um eine allgemeine Anwendung der Lexikographietheorie, die auf eine Verbesserung der Verbreitung des Wissens zielt.

3.2.3 Die Datendistributionsstruktur

3.2.3.1 Verschiedene Texttypen

Ein lexikographischer Prozess für Internetwörterbücher muss eine klare Datendistributionsstruktur aufweisen. Die Planung der Datendistributionsstruktur hat zwei Hauptkomponenten: erstens die Entscheidung über die verschiedenen Wörterbucheinträge, die als Textsortenträger (vgl. Wiegand 1996) funktionieren sollen, und zweitens die Festlegung und Planung der Wörterbuchstrukturen, in denen diese Einträge präsentiert werden sollen und mit denen der Benutzer auf die Daten zugreifen kann. Die Datendistributionsstruktur spiegelt auch die lexikographische Funktion des Wörterbuchs wider.

Kammerer/Wiegand (1998) diskutieren die Rahmenstruktur eines Wörterbuchs. In der Planung der Datendistributionsstruktur ist es wichtig, genau zu entscheiden, wie die Rahmen-

struktur des Wörterbuchs aussehen soll. Hier spielt die Funktion des Wörterbuchs eine wichtige Rolle. Wenn das Wörterbuch z.B. eine kognitive Funktion hat, gilt diese Funktion nicht nur im Wörterverzeichnis, sondern auch in den anderen Texten des Wörterbuchs. In Anlehnung an die Termini *integrierte* und *nicht-integrierte Außentexte* (siehe Bergenholtz, Tarp/Wiegand 1999), wo es sich um Außentexte handelt, die in den genuinen Zweck des Wörterbuchs integriert sind oder nicht, kann man hier von funktions-integrierten Außentexten reden. Dies sind Außentexte, die in eine der Funktionen des Wörterbuchs eingeschlossen sind. Nach Gouws/Steyn (2005) spricht man hier von einer transtextuellen Einstellung gegenüber Wörterbuchfunktionen. In der Datendistributionsstruktur soll der Lexikograph die funktions-integrierten Außentexte planen und genau festlegen, welche Daten in diesen Texten vorkommen sollen.

Jede Sprache hat eine Menge kulturgebundener Ausdrücke. Im Wörterverzeichnis eines einsprachigen oder zweisprachigen Printwörterbuchs gibt es nicht genug Platz für eine lange Erklärung dieser kulturgebundenen Ausdrücke. Eine solche Erklärung kann in einem Außentext gegeben werden. Obwohl ein Internetwörterbuch kein Platzproblem in Wörterbuchartikeln hat, ist eine lange Erklärung der kulturgebundenen Ausdrücke keine Unterabteilung der obligatorischen oder der erweiterten obligatorischen Mikrostruktur. Im Artikel eines Internetwörterbuchs kann eine kurze Beschreibung der kulturgebundenen Ausdrücke gegeben werden, aber dann sollte ein Link den Benutzer auf den Außentext verweisen. Der Außentext unterliegt keinen Platzbeschränkungen und kann eine ausführliche Erklärung der kulturgebundenen Wörter geben – auch in Verbindung mit nicht-lexikographischen Quellen. Die Entscheidung über Außentexte muss während der Formulierung des lexikographischen Prozesses getroffen werden. Die Datendistribution sieht in Internetwörterbüchern anders aus als in Printwörterbüchern.

In den südafrikanischen Bantusprachen gibt es viele kulturgebundene Ausdrücke. In den Nguni-Sprachgemeinschaften, i.e. Zulu, Xhosa, Swati und Ndebele, gibt es z.B. die Tradition der *lobola*. *Lobola* ist der Preis, den ein junger Mann dem Vater seiner zukünftigen Frau bezahlen muss. Doch *Lobola* ist weit mehr als der *Brautpreis*. In einem allgemeinen ein- oder zweisprachigen Xhosa-Wörterbuch gibt es im Artikel der Lemmazeichengestaltangabe *lobola* nicht genug Platz für eine ausführliche Erklärung der Lobolatradition. Zusammen mit anderen kulturgebundenen Wörtern kann *lobola* ausführlich in einem Außentext mit einer kognitiven Funktion erklärt werden. Solche Entscheidungen sollen Teil der Planung der Datendistributionsstruktur sein.

Während der Entwurfsphase soll sich der Lexikograph nicht nur für ein bestimmtes Mikrostrukturprogramm, d.h. die Wahl der verschiedenen Datentypen entscheiden, sondern auch für eine bestimmte Mikrostruktur – oder für mehrere Mikrostrukturen im Fall einer Wörterbuchreihe. In einer datenbasisbasierten Wörterbuchreihe mit zweisprachigen Wörterbüchern für das Sprachenpaar Afrikaans und Deutsch ist es nicht nötig, dass die verschiedenen Wörterbücher alle dieselbe Mikrostruktur aufweisen. Ein Wörterbuch mit Textrezeption als Funktion kann eine rudimentäre Mikrostruktur haben, bei der es keine an Äquivalentangaben adressierten Kotextangaben gibt. In derselben Reihe soll ein Wörterbuch mit Textproduktion als Funktion auch Kotextangaben enthalten, und in diesem Fall braucht der Benutzer eine integrierte Mikrostruktur, in der Kotextangaben jedes Mal in demselben semantischen Subkommentar stehen wie die Äquivalentangabe, an die sie adressiert sind. Visiert das Wörterbuch mit Textproduktion als Funktion z.B. Übersetzer als Benutzer an, brauchen diese vielleicht umfassendere Hilfe bei der Textproduktion. Wo das der Fall ist, kann eine semi-integrierte Mikrostruktur die beste Wahl sein. Für den Benutzer, der nicht viele Kotextanga-

ben braucht, reicht die integrierte Abteilung der Mikrostruktur; aber für den Benutzer, der weitere Hilfe braucht, steht die nicht-integrierte Abteilung zur Verfügung. In diese Abteilung kann der Lexikograph auch verschiedene Kotextdatentypen geben und z.B. Belegbeispielangaben und Kollokationsangaben in unterschiedliche Textblöcke einordnen. Es ist wichtig, die Mikrostrukturen so zu planen, dass der Benutzer einen schnellen und problemlosen Zugriff auf solche Datentypen haben kann. Seit Wiegand (1996) das Konzept der semi-integrierten Mikrostrukturen vorgestellt hat, hat kein Printwörterbuch eine solche Mikrostruktur verwendet. Das ist jedoch eine Art Mikrostruktur, die für Internetwörterbücher sehr geeignet ist. Die Datendistributionsstruktur, die in der Wörterbuchkonzeption eines lexikographischen Prozesses formuliert werden muss, kann also auch zu neuen Mikrostrukturen führen.

3.2.3.2 Einträge

Nach Wiegand (1989) gibt es zwei Arten von Einträgen in einem Wörterbuchartikel, i.e. Angaben und Strukturanzeiger. Die Artikelstruktur eines Wörterbuchartikels ist ein Produkt der Erweiterung oder Anreicherung der Mikrostruktur durch nicht-typographische beziehungsweise typographische Strukturanzeiger. Nicht-typographische Strukturanzeiger führen durch Erweiterung und typographische Strukturanzeiger durch Anreicherung von der Mikrostruktur zur Artikelstruktur. Erweiterung findet man, wo der Artikel neue funktional segmentierbare Einträge enthält, z.B. nicht-typographische Strukturanzeiger. Anreicherung findet man, wo die Änderungen nicht zu funktional segmentierbaren Einträgen führen. Es gibt also keine funktionale Segmentierbarkeit in Wörterbuchartikeln, bei der etwa Belegbeispielangaben in Kursivschrift erscheinen. Obwohl diese Angaben als Belegbeispielangaben identifizierbar sind, weil sie in Kursivschrift gegeben werden, handelt es sich bei den typographischen Anzeigern, i.e. Kursivschrift und Belegbeispielangabe, nicht um zwei verschiedene Einträge. Es liegt also keine Erweiterung, sondern nur eine Anreicherung vor.

Auch in einem elektronischen Wörterbuch verwenden Lexikographen die beiden Typen von Strukturanzeigern. In einem elektronischen Wörterbuch werden häufig auch verschiedene Farben als Strukturanzeiger verwendet. In der Wörterbuchkonzeption muss deshalb die Artikelstruktur erklärt werden, auch unter Hinweis auf die Strukturanzeiger.

Der Zugriff in einem elektronischen Wörterbuch erfolgt nicht immer oder nicht notwendig über die Lemmazeichengestaltangabe als Leitelement des Artikels. Der Benutzer sollte auch einen schnellen, direkten Zugriff auf Angaben haben, ohne die Lemmazeichengestaltangabe als Leitelement zu benutzen. Wo der Zugriff nur artikel-intern erfolgt, müssen die Zugriffsadressen deutlich identifizierbar sein. Der lexikographische Prozess muss hier also den Unterschied zwischen Anreicherung und Erweiterung berücksichtigen.

Obwohl typographische Strukturanzeiger für den erfolgreichen Zugriff auch eine Rolle spielen können, vor allem wenn ein Benutzer den ganzen Artikel rezipiert, sind es doch vor allem nicht-typographische Strukturanzeiger sowie Datenidentifizierungsangaben, die den schnellen Zugriff ermöglichen. Die Möglichkeit des Zugriffs auf bestimmte Datentypen setzt bestimmte Markierungen der einzelnen Datentypen oder Suchbereiche dieser Datentypen voraus. In dieser Hinsicht ist z.B. Zugriff auf Fett- oder Kursivdruck nicht hilfreich. Auf eine angereicherte Artikelstruktur kann also nicht so gut zugegriffen werden wie auf eine erweiterte Artikelstruktur, weil die typographischen Strukturanzeiger nicht funktional segmentierbar und daher in einer Zugriffshandlung, die nicht über die Lemmazeichengestaltangabe als Leitelement des Wörterbuchartikels erfolgt, nicht benutzbar sind. In einer Artikelstruktur, die ein

Produkt einer erweiterten Mikrostruktur ist, ist funktionale Segmentierung möglich, weil diese Erweiterung mit Hilfe von nicht-typographischen Strukturanzeigern, z.B. „*“ oder „\$“, ausgeführt wird. Darum kann der Benutzer über solche nicht-typographischen Strukturanzeiger auf die Daten in einem bestimmten Suchbereich zugreifen.

Aus der Sicht der Informationswissenschaft sind die Strukturanzeiger Teil der Metadaten, weil sie den Benutzer zu den Daten führen. Eine Informationseinheit, die in der Informationswissenschaft auch zu den Metadaten gezählt werden dürfte, sind die Datenidentifizierungsangaben. Sie sind Angaben, weil der Benutzer aus ihnen Informationen über den Wörterbuchgegenstand ableiten kann. So, wie nicht-typographische Strukturanzeiger die verschiedenen Suchzonen identifizieren, identifizieren Datenidentifizierungsangaben die Datentypen der verschiedenen Suchbereiche. Sie sind funktional segmentierbar und deshalb auch Elemente einer schnellen Zugriffsstruktur.

Für den typischen Benutzer, der den ganzen Artikel ansieht, wird die Suche nach einem bestimmten Datentyp vermutlich dann am einfachsten sein, wenn es eine Datenidentifizierungsangabe gibt, die leicht interpretierbar ist, z.B. „SYNONYM“, „KOLLOKATION“ oder „Bedeutung“.

Ein elektronisches Wörterbuch braucht idealerweise eine Artikelstruktur mit einer durch Strukturanzeiger oder Datenidentifizierungsangaben erweiterten Mikrostruktur. Für die Identifizierung der Angabetypen ist die Benennung und Erklärung der verschiedenen Strukturanzeiger schon in der Wörterbuchkonzeptionsphase des lexikographischen Prozesses sehr wichtig. Diese Aufgabe dient im lexikographischen Prozess nicht nur dazu, den Benutzer darüber zu informieren, welche Datentypen im Artikel aufgenommen sind, sondern vor allem, um den elektronischen artikel-internen Zugriff zu ermöglichen.

Dem Benutzer, der den ganzen Artikel sieht, hilft es auch, wenn die Artikelstruktur eine Mikro-Architektur mit einer vertikalen Ordnung in der Suchbereichsstruktur aufweist, wo die verschiedenen Angabetypen in unterschiedliche Textblöcke eingeordnet sind.

In einem lexikographischen Prozess gibt es also viel zu tun, um eine erfolgreiche Datendistributionsstruktur auszuarbeiten.

3.2.4 Die Zugriffsstruktur

Wahrscheinlich der wichtigste Mehrwert eines Internetwörterbuchs sind neue und verbesserte Zugriffsmöglichkeiten. Neben anderen Strukturen soll auch die Zugriffsstruktur ausführlich in der Wörterbuchkonzeption behandelt werden.

Ein Zugriffsmehrwert in Internetwörterbüchern liegt darin, dass der Benutzer nicht immer nur über die Lemmazeichengestaltangabe als Leitelement des Artikels auf eine artikel-interne Angabe zugreifen muss. Wie oben angesprochen, soll der Lexikograph mit Hilfe einer erweiterten Mikrostruktur eine Artikelstruktur entwickeln, die nicht-typographische Strukturanzeiger und Datenidentifizierungsangaben enthält, damit artikel-interner Zugriff möglich wird.

Ein weiterer Vorteil ist, dass die Artikel- und Zugriffsstruktur auch den Zugriff auf nicht-lemmatisch adressierte Angaben ermöglichen kann. Es ist schon hinreichend bekannt, dass

ein Zugriff mit Hilfe eines einzelnen Wortes oder Affixes oder mit einem „search string“ passiert. Diese Zugriffsmöglichkeiten sollen weiter ausgebaut werden.

In der Internetlexikographie brauchen wir mehrere Anwendungen der nicht-lemmatischen Adressierung. In einem Wörterbuchartikel können lemmatisch adressierte Angaben auch sekundäre Leitelemente werden, mit einer eigenen Behandlung, weil es Angaben gibt, die an diese lemmatisch adressierte Angabe adressiert sind. Typisch dafür sind die Kotext- und Kontextangaben in einem zweisprachigen Wörterbuch, die nicht an das Lemmazeichen, sondern an Äquivalentangaben adressiert sind. Eine solche Äquivalentangabe wird also ein sekundäres Leitelement mit einer eigenen Behandlung.

Eine gute Zugriffsstruktur gibt dem Benutzer ggf. auch die Möglichkeit, über ein sekundäres Leitelement auf einen Teilartikel zuzugreifen. Ein Teilartikel ist ein Teil eines Wörterbuchartikels mit einer mikrostrukturellen Angabe als Leitelement und mit einer Behandlung mit Hilfe nicht-lemmatisch adressierter Angaben. Der schnelle Zugriff auf Teilartikel ist ein Merkmal elektronischer Wörterbücher und soll im lexikographischen Prozess festgelegt werden.

Ein Beispiel dieser Zugriffsmöglichkeit findet man in der Darstellung und Behandlung der Kollokationen. Kollokationen sind normalerweise Kotextangaben und verweisen auf typische und frequente Kombinationen des Wortes, das durch die Lemmazeichengestaltangabe repräsentiert wird. In einem allgemeinen Wörterbuch sind Kollokationen keine Lemmakandidaten und darum auch nicht Leitelementträger. Vor allem in einem Wörterbuch mit Textproduktion als Funktion sucht man häufig nicht nur die Kollokation, sondern auch Information über die Kollokation. Der Benutzer braucht Kotextangaben, z.B. Belegbeispielangaben oder weitere Kollokationen mit der spezifischen Kollokation als einer der Konstituenten, die an die Kollokation adressiert sind. Der Benutzer sucht nicht Informationen über das Lemma, sondern nur über die Kollokation. Wenn der Benutzer den ganzen Artikel ansieht, führt ihn eine Datenidentifizierungsangabe wie „KOLLOKATIONEN“ zur relevanten Suchzone. Greift er aber nicht über das Lemmazeichen zu, sondern direkt auf die Kollokation, gibt es zwei Möglichkeiten: Mit Hilfe eines nicht-typographischen Strukturanzeigers oder einer Datentypidentifizierungsangabe kann er auf die Suchzone, die die Kollokationen enthält, zugreifen – so etwa, wenn man z.B. nach „KOLLOKATION *schwören*“ fragt. Der Benutzer hat dann Zugriff auf die bestimmte Suchzone und findet da Kollokationen wie *felsenfest schwören* und *Stein und Bein schwören*. Wenn er aber mehr von einer spezifischen Kollokation wissen will, kann er direkt auf diese Kollokation zugreifen, und dann wird die Kollokation das sekundäre Leitelement eines Teilartikels. Im Artikel des deutschen Wortes *Kritik* findet man Kollokationen wie *scharfe Kritik* und *Kritik üben*. Zugriff auf jede dieser Kollokationen führt den Benutzer zu den jeweiligen Teilartikeln, und in diesen beiden Teilartikeln findet der Benutzer als Teil der Behandlung der spezifischen Kollokation z.B. auch die weitere Kollokation *scharfe Kritik üben*. Mit nicht-lemmatisch adressierten Angaben gibt der Lexikograph dem Benutzer in diesem Teilartikel auch Belegbeispielangaben mit der Kollokation, aber auch weitere Angaben, z.B. Synonyme. In dem Teilartikel der Kollokation *Kritik üben* wird das Synonym *kritisieren* gegeben – und im Artikel des Lemmazeichens *kritisieren* wird dann auch *Kritik üben* als Synonym aufgeführt, siehe Spohr (2005).

Im lexikographischen Prozess muss die Zugriffsstruktur schließlich so geplant werden, dass der Benutzer auch zu Einträgen in den Außentexten gelangt – und in einer datenbasisbasierten Wörterbuchreihe auch zu Einträgen in anderen Wörterbüchern.

3.2.5 Die reflexive Phase

Der lexikographische Prozess eines Internetwörterbuchs braucht auch eine reflexive Phase. Bei einem Printwörterbuch muss der Lexikograph auf eine Revision des Wörterbuchs warten, bevor Fehler korrigiert oder Änderungen gemacht werden können. Internetlexikographie ist ein dynamischer Prozess, der es dem Lexikographen erlaubt, durchgehend Verbesserungen zu machen, ihm aber auch die Gelegenheit bietet, neue Möglichkeiten zu testen. De Schryver/Prinsloo (2000) haben den Terminus *simultaneous feedback* eingeführt – der Lexikograph kann während der Arbeit an einem Wörterbuch eine Rückmeldung von den anvisierten Benutzern bekommen, die ihm bei der weiteren lexikographischen Behandlung helfen kann. Das führt zu einer Art von lexikographischer Demokratie, in der Benutzer Rückmeldungen mit Bezug auf neue Datentypen, neue Strukturen oder die Darstellung der Angaben, der Außentexte, usw. geben. Der lexikographische Prozess sollte daher die Gelegenheit für *simultaneous feedback* vorsehen.

Wichtig ist es auch, dem Benutzer zu vermitteln, dass die Lexikographie dynamisch ist und Wörterbücher nicht notwendigerweise nur von einer kleinen Gruppe Spezialisten gemacht werden.

4. Fazit

Der lexikographische Prozess muss dem Schreiben eines Wörterbuchs vorausgehen. Eine der Herausforderungen der neuen Internetlexikographie ist es, neue Konzepte für die jeweiligen lexikographischen Prozesse zu finden, bei denen aus den alten Prozessen das bewahrt wird, was noch nützlich ist, aber zugleich in innovativer Weise neue Vorschläge gemacht werden, die der lexikographischen Praxis förderlich sind. Und eine verbesserte lexikographische Praxis ist Ausdruck einer verbesserten Metalexikographie.

5. Literatur

- Bergenholtz, Henning/Tarp, Sven/Wiegand, Herbert Ernst (1999): Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. In: Hoffmann, Lothar et al. (Hg.): Fachsprachen. Languages for special purposes. An international handbook of special-language and terminology research. Berlin, S. 1762-1832.
- De Schryver, Gilles-Maurice/Prinsloo, Daniel J. (2000): Dictionary-making process with 'simultaneous feedback' from the target users to the compilers. In: Heid, Ulrich/Evert, Stefan/Lehmann, Egbert/Rohrer, Christian (Hg.): 2000. Proceedings of the Ninth EURALEX International Congress, EURALEX 2000. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, S. 197-209.
- Gouws, Rufus H./Steyn, Mariza (2005): Integrated outer texts: a transtextual approach to lexicographic functions. In: Barz, Irmhild/Bergenholtz, Henning/Korhonen, Jarmo (Hg.): Schreiben, Verstehen, Übersetzen und Lernen: Zu ein- und zweisprachigen Wörterbüchern mit Deutsch. Frankfurt a.M. u.a., S. 127-136.
- Gouws, Rufus H. (2005): Meilensteine auf dem historischen Weg der Metalexikographie. In: Lexicographica 21, S. 158-178.
- Gouws, Rufus H. (2011): Learning, unlearning and innovation in the planning of electronic dictionaries. Fuentres-Oliviera, Pedro A./Bergenholtz, Henning (Hg.): e-lexicography. London, S. 17-29.
- Johnson, Samuel (1747): The plan of a dictionary of the English language. [Facsimile edition, 1970: Menston].
- Johnson, Samuel (1755): A dictionary of the English language. London.
- Kammerer, Matthias/Wiegand, Herbert Ernst (1998): Über die textuelle Rahmenstruktur von Printwörterbüchern: Präzisierungen und weiterführende Überlegungen. In: Lexicographica 14, S. 224-237.
- Landau, Sidney I. (1984): Dictionaries. The art and craft of lexicography. Cambridge.

- McArthur, Tom (1986): *Worlds of reference*. Cambridge.
- Müller-Spitzer, Caroline (2003): Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexikographischen Prozessen. In: *Lexicographica* 19, S. 140-168.
- Prinsloo, Daniel J./Bothma, Theo J.D./Heid, Ulrich/Fass, Gertrud (2012): Devices for information presentation in electronic dictionaries. In: *Lexikos* 22, S. 290-320.
- Spohr, Dennis (2005): *A description logic approach to modelling collocations*. Diplom Thesis: Universität Stuttgart.
- Wiegand, Herbert Ernst (1989): Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. In: *Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie*. 1. Teilbd. (= Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 5.1). Berlin/New York, S. 462-501.
- Wiegand, Herbert Ernst (1996): Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher. In: Wiegand, Hebert Ernst (Hg.). (1996): *Wörterbücher in der Diskussion II*. Tübingen, S. 1-82.
- Wiegand, Herbert Ernst (1998): *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilbd. Berlin/New York.
- Wiegand, Herbert Ernst (2013): Lexikographie und Angewandte Linguistik. In: *Zeitschrift für angewandte Linguistik* 58, S. 13-39.
- Zgusta, Ladislav (1971): *Manual of lexicography*. The Hague/Paris.